The non-classical thesis of emotion

Dylan Evans[1]

Abstract. The non-classical thesis of emotion (NCE) states that the conceptual resources of classical cognitive science cannot adequately account for certain important features of emotion. It also states that these features can be adequately accounted for by employing the conceptual resources of non-classical forms of cognitive science. In this paper I examine one version of this argument — that put forward by Giovanna Colombetti — and argue that she does not succeed in making the case for NCE. In fact, there is a general problem with all forms of NCE, since they all assume that classical cognitive science is too restrictive, when if anything the reverse is true. I conclude by arguing that what is needed here is a theory of pragmatics, rather than a theory of deep conceptual structure. The use of different terms by different groups of cognitive scientists does not grant one group privileged access to cognitive resources inaccessible to the other, but rather serves as a heuristic that directs their attention to different features of the phenomena being studied. Thus the real contribution of non-classical models of emotion is to draw our attention to certain key aspects of emotion requiring explanation that had perhaps been somewhat neglected by classical models.

1 THE NON-CLASSICAL THESIS OF EMOTION (NCE)

During the past decade, some philosophers and psychologists have argued that the conceptual resources of classical cognitive science cannot adequately account for certain important features of emotion. They have further argued that these features can be captured by non-classical forms of cognitive science (eg. [6]). I will refer to such claims as the non-classical thesis of emotion (NCE). In this paper, I examine one particular form of NCE — namely, that put forward by Giovanna Colombetti. Before proceeding to examine Colombetti's arguments in detail, however, let's spell out the main components of NCE in a bit more detail. NCE may be characterised in terms of three lists and three claims, as follows:

Three lists:

1. A list of the conceptual resources of classical cognitive science
2. A list of the conceptual resources of (some form of) non-classical cognitive science, which includes at least some elements not found in List 2.
3. A list of the key aspects of emotion that require explanation

Three claims:

1. The newness of non-classical cognitive science (NNC): Members of list 2 that are not in list 1 cannot be reduced to any combination of the members of list 1
2. The explanatory weakness of classical cognitive science: Some members of list 3 cannot adequately be accounted for by the members of list 1.
3. The explanatory strength of non-classical cognitive science: The same members of list 3 enumerated in claim 2 can be adequately accounted for by the members of list 2 (either on their own, or in addition to the members of list 1).

The three claims are common to all forms of NCE. Differences between the various forms of NCE lie entirely in the different ways that the three lists are populated. None of the proponents of NCE goes so far as to provide exhaustive specifications of all three lists. Not only are their specifications partial, but they are not usually provided in the form of lists at all. Instead, their lists must be reconstructed from hints and ellipses, which makes criticism difficult. Nevertheless, this is what I will attempt do with Colombetti's version of NCE in section 3. Before that, however, I will briefly discuss the various ways in which one might object to NCE.

2 COLOMBETTI'S SPECIFICATION OF CLASSICAL COGNITIVE SCIENCE

Giovanna Colombetti provides a fairly representative example of NCE in a 2003 paper entitled 'Complexity as a new framework for emotion theories' [2]. Colombetti does not talk explicitly about 'classical cognitive science', but she does set out to criticise 'good old fashioned frameworks'. The latter phrase clearly echoes the term GOFAI ('good old fashioned Artificial Intelligence') popularised by the philosopher John Haugeland, and which is synonymous with classical cognitive science [4]. According to Colombetti, the 'good old fashioned frameworks [are] based on *modular* and *hierarchical* perspectives of the mind, which try to explain the elicitation of emotion by positing a *strictly sequential* causal chain of mental and/or physical events' ([2]; emphasis in original). So, here, at least, is a partial specification of list 1 according to Colombetti:

1. Modular processes
2. Hierarchical processes
3. Strictly sequential causal chains of events

Let us examine each of these three terms in detail.

2.1 Modular processes

The term 'module' is used in so many different ways by various sections of the cognitive science community that it often seems to impede communication rather than to facilitate it. In Fodor's classic treatment, a module is a computational mechanism (computational in the sense that it performs transformations on representations) with nine properties, including domain specificity, mandatory operation,

[1] Cork Constraint Computation Centre, Department of Computer Science, University College Cork, Ireland, email: devans@4c.ucc.ie

high speed, and so on [3]. According to Fodor, a mental mechanism must have most or all of these nine properties in order to qualify as a module. Indeed, it is the regular co-occurrence of these nine properties that is supposed to make modularity a robust natural kind. However, as Dominic Murphy and Stephen Stich point out, many evolutionary psychologists tend to use the term 'module' in a much broader and less demanding way, in which the presence of only two of Fodor's nine properties (domain specificity and informational encapsulation) is sufficient for modularity [7]. Others seem to use the term in an even broader way, as referring to any 'subpersonal' mental process — that is, a mental process that has been isolated by some form of functional decomposition — whether or not the process is domain specific or informationally encapsulated. In this very weak sense, a mental process is a module as long as it can be represented as a box in a flowchart model of the mind. I will refer to these three senses of the term 'module' as 'Fodorian modules', 'Darwinian modules', and 'flowchart boxes'.

Colombetti does not spell out how she intends the term 'module' to be understood when she states that 'modular processes' are one of the core conceptual resources of classical cognitive science, but this turns out not to matter, as her argument breaks down no matter how one interprets her use of the term. Either Colombetti is using the term in the most general way, to refer to the boxes in a flowchart model of the mind, or she is using it in a more specific way. On the one hand, if she is using the term in the most general way, then all the apparently new conceptual resources of non-classical cognitive science turn out to be constructed from the familiar building blocks of classical cognitive science, since they can all be expressed in terms of flowcharts. Here, it must be remembered that NCE is a claim about cognitive and explanatory resources, and not just a claim about new models. There is a big difference between the claim that classical cognitive science happened to produce incorrect models of some mental process, and the claim that classical cognitive science lacks the appropriate theoretical resources to produce accurate models of some mental process. Likewise, we should recognise that it is one thing to argue that our model has not decomposed the system in the correct way (that is, it has failed to 'cut nature at its joints' — failed to identify the modules correctly), and quite another to claim that the strategy of functional decomposition is itself flawed. Colombetti represents herself as doing the latter, when in fact she is really doing the former. The new models of emotion she prefers still have functional units — except that she calls them 'micro-components' instead of 'modules'.

On the other hand, if Colombetti is using the term 'module' in a more specific sense, then her specification of list 1 immediately becomes obviously incomplete, since there are many prototypical explanations in classical cognitive science that involve other forms of modularity. If one construes the term 'module' in the full-blown Fodorian sense, for example, then it is clear that classical cognitive science includes other conceptual resources, since many classical explanations do not invoke Fodorian modules. They may involve Darwinian modules or simply flowchart boxes. So we should add these explanatory resources to list 1. But then we are back with the the previous horn of the dilemma, since the inclusion of flowchart boxes in the list of conceptual resources of classical cognitive science undermines the claim about the newness of non-classical science. When we include flowchart boxes and their attendant causal connections in the conceptual resources of classical cognitive science, all the apparently new conceptual resources of non-classical cognitive science turn out to be constructed from the familiar building blocks of classical cognitive science. Conversely, when new models of emotion are provided by non-classical cognitive approaches, the novelty lies in the addition of boxes and arrows to (or, less frequently, the subtraction of boxes and arrows from) earlier classical models — not in the addition of completely new conceptual resources.

So, Colombetti's version of NCE fails in its appeal to the idea of modularity. We may construe Colombetti's use of the term 'module' in a general way or a specific way, but each of these options is problematic. Colombetti's argument is thus caught on the horns of a dilemma.

2.2 Hierarchical processes

A similar argument may be made with respect to the second of the conceptual resources that Colombetti attributes to the armoury of classical cognitive science — hierarchical processes — since this phrase is also ambiguous. A set of cognitive processes or modules could be hierarchical in several different ways. For example, the hierarchy could be coordinative (in which one process modulates another) or competitive (in which one process subsumes or replaces another).

If the hierarchy in question is coordinative, then a hierarchical model is one in which one process modulates, but is not modulated by, another. Some classical models of emotion have been proposed which are hierarchical in this sense - but if such models are wrong, this is a fault of the model, and not the of the classical approach in general, since non-hierarchical models can also be constructed from the resources of classical cognitive science. The argument is similar to that concerning sequential verus circular causal chains (see Sections 2.3 and 3.2 below).

If the hierarchy in question is competitive, then a hierarchical model is one which processes are layered, as in a subsumption architecture, so that when the input conditions of more than one module are met, then the highest level module gets to control motor processes. In this case, the layering could be fixed, or it could be variable. With fixed layering, module 1 always takes precedence over module 2, for example. With variable-order layering, which module takes precedence might depend on the relative values of a variable in each module. One can imagine many other ways in which a set of modules can be hierarchical too. For example, a system could be termed hierarchical if there is at least one 'information bottleneck' through which all information must pass on its way from sensors to effectors, one central point at which any conflicts are resolved and decisions taken. The only thing all these meanings have in common is that they are serial systems, since hierarchies always specify an ordering of some kind. Only a completely parallel system would not be hierarchical.

Either Colombetti means hierarchical in one of the many specific senses of the term, or she means it in the most general sense, as excluding only entirely parallel systems. On the one hand, if she is using the term in the most general way, then all the apparently new conceptual resources of non-classical cognitive science turn out to be constructed from the familiar building blocks of classical cognitive science, since they can all be expressed in terms of serial systems of one kind or another. This applies to parallel machines too, since any parallel machine can be simulated on a serial machine. On the other hand, if she is using the term in a more specific sense, then her specification of list 1 immediately becomes obviously incomplete, since there are many prototypical explanations in classical cognitive science that involve other forms of hierarchy. So we should add these explanatory resources to list 1. But then we are back with the the previous horn of the dilemma, since the inclusion of all types of hierarchical process in the list of conceptual resources of classical cogni-

tive science undermines the claim about the newness of non-classical science.

2.3 Strictly sequential causal chains of events

The third of the conceptual resources that Colombetti attributes to classical cognitive science is the concept of 'strictly sequential' causal chains. It is not clear what subset of causal chains Colombetti wishes to pick out by the use of the phrase 'strictly sequential'. Does she mean causal chains that do not allow bifurcation or recursion?

If not, then classical cognitive science can account for arbitrarily complex causal chains, and hence the supposed novelty of non-classical cognitive science cannot be due to anything distinctive about the kind of causal processes it invokes. If she does, then her specification of list 1 is obviously incomplete, since there are many examples of models in classical cognitive science that involve bifurcating and/or recursive causal chains. Either way, then, Colombetti's version of NCE fails in its appeal to the idea of 'strictly sequential' causal chains. This point is explored further in Section 3.2 below.

3 COLOMBETTI'S SPECIFICATION OF NON-CLASSICAL COGNITIVE SCIENCE

Colombetti does not use the term 'non-classical cognitive science', but she does talk about 'the dynamical systems approach in cognitive science' (or the 'dynamical perspective'), and explicitly contrasts this approach with 'the good old fashioned frameworks' that we have already claimed to be coterminous with classical cognitive science. This is broadly in line with most of the claims made on behalf of non-classical cognitive science, which tend to focus on the 'dynamical hypothesis in cognitive science' [8], or on strong claims about the embodiment and/or situatedness of cognition [1], or both. Thus we may take her characterisation of the dynamical systems approach to be her specification of non-classical cognitive science.

Colombetti does not provide a formal list of the conceptual resources of the dynamical systems approach, but she does mention the following two ideas as being distinctive features of this approach:

1. Collective action of micro-components
2. Circular causation

The first of these is explicitly contrasted with the 'hierarchical' processes supposedly invoked by classical cognitive science, and the second with the 'strictly sequential causal chains' that classical cognitive science is supposedly restricted to. Since these terms are problematic, as has already been noted in Sections 2.2 and 2.3, it is hardly surprising that their supposed opposites are similarly problematic.

3.1 The collective action of micro-components

A micro-component is just another name for a kind of module (module in the broadest sense of the term). In other words, if a system can be analysed as a collection of micro-components, then the strategy of functional decomposition works.

The term 'collective action' is not well defined by Colombetti, but it is clear from the context that she intends it to be the contrary of the term 'hierarchical'. But as we have already seen in Section 2.2, this term is not well defined either. If the term 'hierarchical' is used in an overly specific way, then 'collective action' embraces a whole variety of processes that are clearly among the conceptual resources of classical cognitive science. Alternatively, if the term 'hierarchical' is used in the most general sense of the term, then there is no model of collective action that cannot be translated into some hierarchical model, and so this cannot be the basis of any claim of superiority for non-classical cognitive science.

3.2 Circular causation

The only clear interpretation of the phrase 'circular causation' is a bidirectional or reciprocal causal connection — in other words, a feedback loop. Feedback loops are an elementary resource of classical cognitive science, so for NCE to succeed, Colombetti must claim that there is more to the idea of circular causation than just feedback. This is indeed what she does — but she fails to say anywhere what the supposed extra content is.

Colombetti dismisses many of the appraisal theories that are often held to typify classical cognitive theories of emotion, which model the relationship between appraisal processes and emotions in the following way (see fig 1):

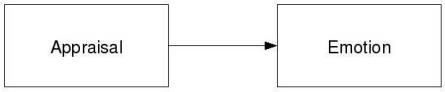

Figure 1. A typical appraisal theory view of the relationship between appraisal and emotion

In this diagram, the box marked 'appraisal' designates some selection mechanism that takes perceptual input (after very little or a great deal of pre-processing, depending on the flavour of the appraisal theory) and outputs an emotion descriptor. The box marked 'emotion' takes an emotion descriptor as input and transforms this input into various output signals, such as control variables for modulating decision-making processes and physiological variables. Colombetti appeals to Lewis, who criticises such 'one-way causal explanations' on the grounds that they do not take into account the fact that emotions are not just caused by appraisal processes, but also affect those appraisal processes [6]. This would seem to be an objection to the particular model in Figure 1 rather than to the conceptual resources of classical cognitive science per se, since we can amend the model simply by adding another arrow — one running from the emotion module back to the appraisal module. The new arrow indicates that among the various output signals from the emotion module is one that modulates the appraisal process. For example, if the emotion module receives the descriptor 'fear' as input from the appraisal module, it might feed back a signal to the appraisal module that makes it more sensitive to perceptual input that has triggered fear in the past. This would be a case of positive feedback, since the organism would be more likely to appraise a situation as fearful when in a state of fear, and so emotional states would have a kind of inertia.

The addition of a second arrow to Figure 1 would clearly transform it into a different model of the appraisal-emotion process, but it would not involve any new cognitive resources. The explanatory resources it employs are just those of classical cognitive science: modular decomposition and causal connections. The new model includes a feedback loop, while the former did not, but while feedback loops have very different dynamics to simple feedforward mechanisms, it does not require any new explanatory resources to account for them.

This may seem too obvious to be worth stating, but it is ignored or denied by such eminent proponents of NCE as Lewis. The idea that the existence of "continuous reciprocal causation" (that is, feedback) renders a system resistant to modular decomposition is quite ridiculous. On the contrary: one can only analyse the reciprocal causation of two subsystems if one has taken prior advantage of the strategy of modular decomposition. In other words, the notion of reciprocal causation, far from rendering the practice of modular decomposition redundant, actually requires it.

4 CONCEPTUAL RESOURCES, COGNITIVE ARCHITECTURES, AND COGNITIVE MODELS

It is important to note that NCE is not a claim about the existence of new models or theories of emotion, but a claim about conceptual resources (or, more precisely, a claim about the relationship between models and conceptual resources). If a new model or theory of emotion accounts for hitherto refractory aspects of emotional phenomena, but can be entirely explicated by recourse to the conceptual resources of classical cognitive science, then the existence of the new model provides no support to NCE. This is often overlooked by proponents of NCE. Typically, supporters of NCE get excited about a new model of emotion that is expressed in terms that are not part of the conceptual toolbox of classical cognitive science. The fact that the new model accounts for aspects of emotion that have previously been neglected by models developed in the classical idiom is then taken to show that the classical idiom is incomplete. But this is a non-sequitur, for it neglects the possibility that the new model can also be expressed in terms that are drawn entirely from the the classical idiom. For example, Marc Lewis prefers models of emotion that are couched in terms of linked sets of dynamical equations over the traditional block diagrams of classical cognitive science, since the former'highlight reciprocal, multiple, and recursive causality in both the emergence and stabilization of emotionappraisal states' [6, p. 193]. He fails to see that these are not mutually exclusive; a model expressed in terms of a set of linked equations can always be translated into a suitably annotated block diagram.

It is important to note that the reverse is not always true. A block diagram cannot always be translated into a set of linked equations. If, for example, it is not properly annotated and/or fails to meet certain well-formedness criteria, then a block diagram will not be specified precisely enough to be translated into a single set of equations. This might mean simply that the block diagram corresponds to a variety of sets of equations: it is simply not yet specified precisely enough for us to determine which set of equations it expresses. Alternatively, it could mean that the block diagram does not map onto a set of equations at all, since it is not a numerical model; the arrows connecting the boxes might represent the transmission, not of some mathematical quantity, but of some syntactic representation, as in a state diagram that represents a finite sate machine.

Since models expressed in terms of sets of linked equations can always be translated into block diagrams, but not always vice versa, it follows that block diagrams comprise a more powerful set of conceptual resources than sets of linked equations. Block diagrams are more powerful than sets of linked equations in another sense too; they provide a clear model of the internal components of the mechanism that produces the behaviour, while the linked equations merely describe the nature of the observable behaviour. If block diagrams are the canonical form of models in classical cognitive science, and sets of linked equations are the canonical form of models in dynamical cognitive science, as Lewis claims, it follows that the conceptual resources of classical cognitive science are more powerful than the conceptual resources of dynamical cognitive science.

This general point undermines all the various forms of NCE. All forms of NCE require the combined conceptual resources of classical cognitive science and non-classical cognitive science to be greater than the conceptual resources of classical cognitive science alone. More formally, if C is the set of conceptual resources of classical cognitive science, and N is the set of resources for non-classical cognitive science, then if NCE is true, the relative complement of N relative to C must be non-empty.

However, the problem with classical cognitive science, if there is one, is that its conceptual resources are all-encompassing. As a theory, it is not constrained enough. Take the Soar cognitive architecture developed by John Laird, Allen Newell and Paul Rosenbloom, for example [5]. Soar is about as good an example of classical cognitive science as anyone could hope for. It was designed as a common format for expressing a whole variety of cognitive models. Yet Soar is Turing-complete, so it can be programmed to represent any kind of computational cognitive model at all. So, for most of its critics, the problem with Soar is not that it is too constrained, but that it does not embody enough constraints to act as a good psychological theory.

The reference to Soar is particularly apt, since the current discussion would be better understood by cognitive scientists themselves (rather than by the philosophers of cognitive science who tend to dominate the discussion) if it were couched in the terminology of'cognitive architectures' rather than that of'conceptual resources'. A cognitive architecture is, in fact, a specification of the kind of conceptual resources that may be used to construct a set of consistent cognitive models. Classical cognitive science is perhaps best seen as a set of cognitive architectures (comprising Soar, ACT-R, and others), while non-classical cognitive science is a different set (comprising subsumption architectures, neural networks, dynamical models, among others). For any pair of architectures, and any cognitive model, the cognitive model can always be programmed in both, or just the classical architecture — but never in the non-classical architecture alone.

5 CONCLUSION

The classical and non-classical forms of cognitive science certainly sound different. The key terms of the latter are rarely, if ever, to be found in the former. However, these terminological differences do not reflect any deep conceptual rift, since there is nothing in non-classical explanations that cannot be translated into the terms of classical cognitive science. Yet the different terminology employed by classical and non-classical forms of cognitive science does make a difference to the way that the proponents of each go about their research. Terms like'circular causation' summon up in the researchers' mind a set of studies that have put special emphasis on feedback loops (even though the researcher might explicitly state that they are concerned with something'more' than mere feedback) and so perhaps lead the researcher who'feels at home with' this terminology to discover feedback loops that he or she might might otherwise have missed.

What is needed here is a theory of pragmatics, rather than a theory of deep conceptual structure. The use of different terms by different groups of cognitive scientists does not grant one group priveleged access to cognitive resources inaccessible to the other, but rather serves as a heuristic that directs their attention to different features of the phenomena being studied.

ACKNOWLEDGEMENTS

The research for this paper was supported by Marie Curie Transfer of Knowledge Action no. MTKD-CT-2006-042563. Thanks are also due to Ric Wallace for his comments on an earlier draft of this paper.

REFERENCES

[1] A. Clark, *Being There: Putting Brain, Body, and World Together Again*, MIT Press, Cambridge, Mass., 1997.
[2] G. Colombetti, 'Complexity as a new framework for emotion theories', *Logic and Philosophy of Science*, **1**, (2003).
[3] J. A. Fodor, *The Modularity of Mind*, MIT Press, Cambridge, MA, 1983.
[4] J. Haugeland, 'What is mind design?', in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, ed., J. Haugeland, MIT Press, Cambridge, Mass. and London, England, (1996).
[5] John E. Laird, Allen Newell, and Paul S. Rosenbloom, 'Soar: an architecture for general intelligence', *Artif. Intell.*, **33**(1), 1–64, (September 1987).
[6] M. D. Lewis, 'Bridging emotion theory and neurobiology through dynamic systems modeling (with commentary)', *Behav Brain Sci*, **28**(2), (April 2005).
[7] D. Murphy and S. Stich, 'Darwin in the madhouse: evolutionary psychology and the classification of mental disorders', in *Evolution and the Human Mind : Modularity, Language and Meta-Cognition*, eds., P. Carruthers and A. Chamberlain, Cambridge University Press, Cambridge, (2000).
[8] T. van Gelder, 'The dynamical hypothesis in cognitive science.', *Behav Brain Sci*, **21**(5), (October 1998).