Classicalism and Cognitive Architecture

Tim van Gelder
Research School of Social Sciences
Australian National University,
Canberra ACT 0200, Australia
tvg@coombs.anu.edu.au

Lars Niklasson
Department of Computer Science,
University of Skövde, S-54128 Sweden
lars@ida.his.se

Abstract

This paper challenges the widely accepted claim that "classical" cognitive architectures can explain the systematicity of cognition (Fodor & Pylyshyn, 1988). There are plausible ways of rendering more precise the systematicity hypothesis (as standardly formulated) in which it *is* entailed by classical architectures, and other plausible ways in which it is *not*. Therefore, it is not a determinate issue whether systematicity is entailed, and hence explained, by classical architectures. The general argument is illustrated in a particular domain, the systematicity of deductive inference. In the case of the capacity to carry out the inference *modus tollens*, the systematicity hypothesis can be made precise in two ways, one entailed by classical architectures, another which is not. Further, the latter, but not the former, accurately describes the actual empirical phenomenon. Put another way, the clumps that these deductive inference capacities come in are not the clumps that are entailed by classical architectures. Therefore, in this area at least, systematicity considerations count *against* the classical conception of cognitive architecture.

In their well-known paper Connectionism and Cognitive Architecture, Fodor & Pylyshyn (1988) argued that connectionism cannot constitute a viable alternative to the so-called "classical" conception of cognitive architecture, on the grounds that, unlike classical architectures, it cannot provide any explanation of the pervasive empirical phenomenon of systematicity. Therefore, the only proper role for connectionism is to investigate how classical architectures might be implemented. This argument—described by Pylyshyn as one of only two major arguments for the classical approach (Pylyshyn, 1989)—has prompted much connectionist modeling and philosophical debate. The main point of contention has been whether connectionism can deliver an adequate non-classical explanation of systematicity. Most contributors to these debates have paid little or no attention to the alleged empirical phenomenon of systematicity itself, and to whether classical architectures can in fact explain systematicity.

Here we take a very different approach. We will argue that, according to Fodor and Pylyshyn's own standards of explanation, classical architectures presently cannot be claimed to explain systematicity, and further, that there is reason to believe that classical architectures *cannot* explain systematicity. Consequently, there can be no argument from any alleged failure of connectionism to explain systematicity to the superiority of the classical approach. Indeed, if anything, systematicity considerations currently count as an argument *against* the classical conception.

What Is It To Explain Systematicity?

Fodor & Pylyshyn place a strict constraint on what counts as genuine explanation. For an hypothesis H to explain some phenomenon S, H *alone* must *entail* S. This constraint is essential to their dismissal of connectionism. They imagine a defender of connectionism building systematicity into a particular connectionist model and claiming, on this basis, that connectionism can indeed explain systematicity. Fodor & Pylyshyn reply that this would be insufficient. Since *all* natural cognitive systems are systematic, systematicity must *follow from the basic nature of architecture alone*, and not merely be compatible with the architecture (p.50).

Consequently, if the *classical* conception of cognitive architecture explains systematicity, it likewise must entail systematicity from the basic nature of the architecture alone. The classical conception, expressed as an empirical hypothesis, is

> *CA*: All natural cognitive systems contain (a) mental representations with combinatorial constituent structure and compositional semantics, and (b) mental processes that are sensitive to the combinatorial structure of the representations.

This hypothesis must *entail* the alleged empirical phenomenon, which is that

> *S*: All natural cognitive systems are systematic.

This entailment is neither immediate nor obvious. Whether it goes through at all depends, among other things, on what the empirical phenomenon S actually is, and this in turn depends on what the *concept* of systematicity is. *Until systematicity is adequately clarified, we cannot know whether classical architectures explain it.*

What Is Systematicity?

Systematicity appears nowhere in the cognitive science literature before Connectionism and Cognitive Architecture: not as an empirical hypothesis, nor as a concept; not even the term.[1] It is therefore surprising

[1] *Productivity*, of course, occurs frequently in the literature, and productivity is one component of systematicity. However, a concept does not exist merely because one component of it exists. Note, moreover, that Fodor & Pylyshyn explicitly decline to rely on productivity for the force of their argument; it is only the *other* components of systematicity that matter.

that Fodor & Pylyshyn give no clear, succinct and precise definition of the concept, or description of the empirical phenomenon. Careful extrapolation from the scattered hints and definitional tidbits they do provide leads to the following empirical hypothesis:

> *Systematicity according to Fodor & Pylyshyn* (*SFP)*: All cognitive systems (humans and other animals) are systematic, i.e., are such that their ability to do some things of a given cognitive type (including at least "thinking a thought" and making an inference) is intrinsically connected with their ability to do other, structurally related things of that type.

This statement is as precise as Fodor & Pylyshyn get. However, it is not sufficiently precise for determining whether classical architectures explain systematicity. There are ways of sharpening *SFP* such that it *is* entailed by *CA*, and ways of sharpening it such that it is *not* entailed by *CA*. Therefore, it is not a determinate issue whether *CA* as it stands entails *SFP*, or systematicity in general.

The remainder of this section states this argument in a little more detail.

How might *SFP* be rendered precise? As noted, we cannot turn to the existing literature for help, for there is none bearing directly on the issue. (This may explain the otherwise curious fact that Fodor & Pylyshyn, in the body of Connectionism and Cognitive Architecture, cite *no empirical literature whatsoever* in support of their claim that cognition is systematic.) Consequently, we must begin from scratch.

The basic idea behind systematicity is that any organism able to do one thing of a given type is able to do other, structurally related, things of that type. To render systematicity precise is to get clear on what cognitive performance types there are, and on what other things of a given type an organism would have to be able to do, if it can do some particular thing of that type. This suggests the following schema:

> *Systematicity Schema*: For every organism O, and any given cognitive performance t of type T, there is some set $M_{O,t}$ of "structurally related" performances such that O is capable of all and only the performances in $M_{O,t}$.

This schema would then be fleshed out for particular organisms and performance types in order to provide empirically applicable tests.

There are many ways the Systematicity Schema might be filled out in detail for particular performance types. Some of these ways are entailed by the hypothesis that cognitive architectures are classical (*CA*). Some are not. Suppose, for example, that the sets $M_{O,t}$ were specified to be some proper, non-arbitrary subsets of the sets that would be entailed by classical architectures. (An example is given in the next section.) Then those cognitive capacities would clearly be systematic, but systematicity of this kind would *not* be entailed by classical architectures. for classical architectures entail that the organism is capable of performances that are *not* in those sets. Consequently, *in the absence of any particular specification of how the Systematicity Schema is to be filled out, there is no determinate answer to whether classical architectures entail and hence explain systematicity*. Further, since no-one has in fact provided such specification in any reasonable detail, no-one can presently justifiably assert that classical architectures *do* entail and hence explain systematicity.

Think of it this way. *CA* entails that cognitive capacities come in clumps, i.e., are systematic. *SFP* asserts that cognitive capacities come in clumps. Does *CA* therefore entail *SFP*? Not necessarily. It only does so if the *kind* of clumps entailed by *CA* are the *same kind* of clumps picked out by *SFP*. So, what kind of clumps does *SFP* pick out? Well, that's not clear at all. On one way of reading SFP, the clumps are the same. On another way, the clumps are not. Until we've settled on a specific way of reading *SFP*, we just can't say whether the entailment is there. Unfortunately, Fodor & Pylyshyn didn't provide any specific reading, and nobody else has either. So, we can't now say that classical architectures do entail, and hence explain, systematicity.

How Classical Architectures Do *Not* Explain Systematicity

The previous section argued that it cannot now be said that classical architectures entail systematicity. This section argues for a different point: that in one domain at least, classical architectures *do not* entail the actual empirical facts of systematicity. In making this second argument, we provide a concrete illustration of the key premise of the first argument, which is that there are various ways of making systematicity precise, some which are, and some which are not, entailed by classical architectures.

Fodor & Pylyshyn single out the *systematicity of inference* as a key component of the wider phenomenon of systematicity. It is, roughly, the idea that the ability to make some inferences is intrinsically connected to the ability to make other, logically related inferences. They offer no precise definition of the phenomenon and cite no literature in its support, but do anecdotally illustrate what they have in mind the following way:

> You don't, for example, find minds that are prepared to infer John went to the store from John and Mary and Susan and Sally went to the store and John and Mary went to the store but not from John and Mary and Susan went to the store. (p.48)

Perhaps; perhaps not. In any case, the import of such an casual observation for the general phenomenon of systematicity of inference is entirely unclear.

A more appropriate procedure is to fill out the Systematicity Schema for particular inference types. Here we discuss only one, *modus tollens* ($A \supset B$, $\sim B \Rightarrow \sim A$). Note that *modus tollens* is one of the simplest and most common of all inference types. If classical architectures fail to explain the actual systematicity of *modus tollens*, this will significantly undermine the claim that classical architectures can in fact explain the systematicity of inference and indeed systematicity in general.

There are at least two ways to flesh out the Systematicity Schema for this form of conditional inference. One is such that the systematicity of this capacity is entailed by classical architectures. Such architectures postulate mental processes that operate on mental representations in a way that respects their combinatorial structure. Classical architectures therefore predict that (subject to resource constraints) any cognitive system that can perform *any* instance of *modus tollens* (i.e., can construct conditionals and negations, and is able to draw the appropriate conclusion) will be able to perform *every* instance, since all such instances have the same combinatorial structure. In particular, classical architectures predict that, since mental processes are sensitive to structure, such features of the inference instance as the *content* of the constituent symbols or their *frequency of prior occurrence* should be irrelevant, since such features make no difference to the combinatorial structure.

More precisely, classical architectures entail the following systematicity sub-hypotheses. Let $M_{O,MT}$ be the set of inferences by substituting into the *modus tollens* schema $(A \supset B, {\sim}B \Rightarrow {\sim}A)$ any symbol in the set of symbols available to O. Then we have:

> Systematicity of Modus tollens (SMT): Any organism O capable of performing any instance of $M_{O,MT}$ is capable of performing every instance of that set.

(Note that this instantiation of the Systematicity Schema has been simplified by assuming that the set M does not depend on any particular performance t.) Hypothesis SMT expresses in precise terms one way in which cognitive capacities can be said to come in clumps. It is entailed by the hypothesis that cognitive architectures are classical in form. Does it accurately describe the kind of clumps that cognitive capacities actually come in?

Conditional inference has been the target of much psychological investigation. Though this investigation was not specifically directed at evaluating any systematicity hypothesis, it does shed a certain amount of light on the issue. The general situation is dramatically illustrated by Table 1, from a study by Kern, Mirels & Hinshaw (1983).

In this study, scientists were presented with conditional inferences of four kinds and asked whether certain conclusions followed. Some were presented inferences in abstract form (Do $P \supset Q$ and *not-Q* imply *not-P*?) and others in concrete form (Do *If Rex is a terrier, then he likes apples*, and *Rex does not like apples*, imply *Rex is not a terrier*?). For current purposes, the crucial thing to notice is the *disparity* in performance between abstract and concrete instances. Note, for example, only 41% of scientists in one group correctly recognized the validity of *modus tollens* in an abstract case, whereas 69% in another recognized its validity in a concrete case. Assuming representativeness, this suggests that around 30% of people do not perform identically on structurally identical inferences, i.e., directly violate hypothesis SMT. The clumps that these people's cognitive capacities come in are *not* the clumps entailed by the hypothesis that they have a classical architecture.

For various reasons it would be inappropriate to place too much weight on these figures alone. The moral they suggest has, however, been borne out repeatedly in numerous systematic studies. The difference between the abstract and concrete cases is that the latter have meaningful content. The critical role of content in conditional inference has been confirmed repeatedly in one of the most-studied tasks in the psychology of inference, the Wason card selection task (Wason, 1966). In the standard version of this task, four cards are laid out on a table before the

Table 1. Scientist's performance on simple conditional inferences. From Kern, Mirels & Hinshaw (1983)

recognized validity (or invalidity) of:	Referents	Discipline			Overall % correct, collapsed across disciplines	subjects with some logic training (n=26)
		Psychologists	Biologists	Physicists		
modus	abstract	83	92	92	89	100
ponens	concrete	100	100	100	100	100
denial of	abstract	83	58	75	72	67
antecedent	concrete	75	75	100	83	100
affirmation of	abstract	83	75	75	73	87
consequent	concrete	75	50	92	72	82
modus	abstract	33	33	58	41	53
tollens	concrete	50	83	75	69	91

subject; each card has a number on one side and a letter on the other. The subject is asked which cards would logically need to be turned over in order to find out whether a rule is true or false. The rule is a conditional: if there is an A on one side, there is a 3 on the other.

Famously, subjects generally perform very badly. The important point here, however, is that (a) selecting the fourth card corresponds to correct performance of *modus tollens*; (b) subjects very rarely perform this inference correctly in the standard version of the task; and (c) numerous studies have shown that performance is often much better if the cards and rule are changed to something familiar, such as the rule "If a person is drinking beer, then that person must be over 19 years of age." (Griggs & Cox, 1982). For an overview of these studies, see Evans (1989) Ch.4.

Results such as these have led most psychologists of inference to conclude that no account of the empirical facts of human inference will succeed if it attempts to explain inferential capacities in terms of the *formal structure* of inferences alone - i.e., in terms only of the basic resources offered by classical architectures. The following observation by two of the foremost specialists in the psychology of deductive reasoning, Wason and Johnson-Laird, is typical:

> The emphasis which we have placed on the importance of content in reasoning shows that a purely formal, or syntactic, approach to it may suffer from severe limitations...But one can turn this argument round, and examine the usefulness of formal logic in constructing psychological models of reasoning. For some considerable time we cherished the illusion that this was the way to proceed, and that only structural characteristics of the problem mattered. Only gradually did we realize first that there was no existing formal calculus which correctly modeled our subject's inference, and second that no purely formal calculus would succeed. (1972 pp. 244-5)

Compare this with Fodor & Pylyshyn's remarkable, unsupported assertion that: "It's a 'logical' principle that conjunctions entail their constituents (so the argument from *P&Q* to *P* and to *Q* is valid). Correspondingly, its a psychological law that thoughts that *P&Q* tend to cause thoughts that *P* and thoughts that *Q*, all else being equal." (p.46)

There is, of course, an alternative way of filling out the Systematicity Schema for *modus tollens*, for anyone who has the patience. We just transcribe the actual data, painstakingly gathered by experimental psychologists, into a new Systematicity of *Modus tollens* hypothesis, *SMT2*. This alternative hypothesis describes another way cognitive capacities come in clumps. Moreover, it is trivially guaranteed to describe the actual empirical phenomena to the best of our current knowledge. However, *SMT2* is *not* entailed by the hypothesis that all cognitive architectures are classical. The classical conception of cognitive architecture entails that all subjects will perform the same on structurally identical inferences, regardless of the content of the symbols. *SMT2* asserts that people do not perform the same on identically structured inferences.

In short, there are at least two different ways of rendering systematicity (*SFP*) precise in the case of one cognitive capacity, the capacity to correctly make *modus tollens* inferences. One (*SMT*) is entailed by the classical approach; another (*SMT2*) is not. Therefore, classical architectures cannot be said to entail the vague, general systematicity hypothesis *SFP*. Further, it is *SMT2*, not *SMT*, which correctly describes human performance. Therefore, classical architectures fail to explain the systematicity of *modus tollens* capacities, such as it is. Thus, in this area at least, systematicity considerations count as an argument *against* the classical approach.

Discussion

Enthusiasts of classical architectures might respond by qualifying their explanatory ambitions. They might abandon their empirical hypothesis that inference is systematic in just the way entailed, and hence explained, by classical architectures, and argue only that classical architectures are to be preferred to connectionist ones on the grounds that "thought" or perhaps language is systematic, and only classical architectures explain this. However, apart from being a major concession, this would be a rather *ad hoc* maneuver. If cognitive architectures are classical in basic form, why are only certain cognitive capacities, and not others, systematic in the way entailed by such architectures? Indeed, we can ask why Fodor & Pylyshyn did not assert as empirical fact that *all* cognitive capacities are systematic, since that is what is entailed by the hypothesis that cognitive architectures are classical. Why did they carefully select out certain cognitive capacities ("thought", inference), assert that *those* are systematic, and that classical architectures are to be preferred because they explain those capacities?

An alternative response is that what classical architectures really entail, and hence explain, is our *competence* in deductive inference; the actual details of human performance arise from an interaction between a pristine competence and various kinds of contingent limitations, e.g., in resources. Thus the actual facts of systematicity do not automatically rule out the possibility that cognitive architectures are classical. To this, three responses. First, the argument of this paper has not been that the actual empirical facts of systematicity *rule out* the classical hypothesis; merely that classicalists cannot currently claim empirical virtue in being able to explain systematicity. Second, note that in Connectionism and Cognitive Architecture, Fodor & Pylyshyn *explicitly declined* to rely on competence/ performance distinctions, acknowledging that the hypothesis of ideal competence was held in suspicion in the opposing camp (connectionists). Third, explanations of facts of performance that begin from a competence and degrade that competence by means of contingent limitations must be able to give *plausible* explanations of *how* those contingent limitations give rise to the observed performance. In the case of deductive inference, there is no obvious plausible way to do this. Classicalists are welcome to take up the challenge.

What kind of general conception of cognitive architecture would stand a realistic chance of entailing, and hence explaining, the actual facts of systematicity? One lesson to be learned from the brief discussion of deductive inference is that the actual clumpiness of real cognitive performance must reflect at least two factors. Obviously, structural relationships are very important. The structural similarities in the tasks themselves, and presumably in the internal states that make possible performance of those tasks, have much to do with the fact that our cognitive capacities come in clumps. However, this cannot be the whole story, since our cognitive capacities apparently do not always come in the kind of clumps that are entailed by such considerations. The critical role of content in deductive inference capacities suggests *familiarity* also has much to do with the actual shape of the clumps that cognitive capacities come in. Therefore, it is plausible that any conception of cognitive architecture that built in, from the outset, *both* structured internal states and structure-sensitive processes *and* a sensitivity to experience would be able to deliver an explanation of the actual facts of systematicity. Interestingly, this vague description embraces the kind of architecture that many connectionists work with (e.g., Elman, 1991).

It is an open question whether the basic principles defining this sub-class of connectionist architectures can in fact be formulated in such a way as to entail, and hence explain, the basic outline of the clumps that cognitive capacities actually come in. The point of this paper has been to challenge the nearly universal acquiescence in Fodor and Pylyshyn's claim that classical architectures entail and hence explain systematicity. This acquiescence has been made possible by a widespread reluctance to engage in the kind of tedious empirical inquiries that would be required to ascertain the actual nature of the clumps that cognitive capacities come in.

Acknowledgements

Thanks to Leah Savion, Frank Jackson and Robert Hadley for useful discussions and comments. This research was supported in part by an award from the Australian Research Council to the first author and from the University of Skövde to the second author.

References

Elman, J. L. (1991) Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, **7**, 195-225.

Evans, J. S. B. T. (1989) *Bias in human reasoning : causes and consequences*. Hillsdale N.J.: Lawrence Erlbaum Associations.

Fodor, J. A., & Pylyshyn, Z. (1988) Connectionism and cognitive architecture: a critical analysis. *Cognition*, **28**, 3-71.

Griggs, R. A., & Cox, J. R. (1982) The elusive thematic materials effect in the Wason selection task. *British Journal of Psychology*, **73**, 407-420.

Kern, L. H., Mirels, H. L., & Hinshaw, V. G. (1983) Scientists' understanding of propositional logic: an experimental investigation. *Social Studies of Science*, **13**, 131-46.

Pylyshyn, Z. (1989) Computing in Cognitive Science. In M. A. Posner ed., *Foundations of Cognitive Science*. Cambridge MA: MIT Press.

Wason, P. C. (1966) Reasoning. In B. M. Foss ed., *New Horizons in Psychology I*. Harmandsworth: Penguin.

Wason, P. C., & Johnson-Laird, P. N. (1972) *Psychology of Reasoning: Structure and Content*. Cambridge MA: Harvard University Press.