Can double dissociation uncover the modularity of cognitive processes?

Giorgio Ganis
Department of Cognitive Science
University of California at San Diego
9500 Gilman Dr.
92093-0515, La Jolla, CA.

Nick Chater
Department of Psychology
University of Edinburgh
7, George Square,
Edinburgh, EH8 9JZ, U.K

Abstract

Neuropsychological evidence has proved influential both in testing pre-existing cognitive theories and in developing new accounts. It has been argued that dissociations, and, in particular, double dissociation are particularly valuable in developing new theoretical accounts, since they may reveal the gross structure or "modularity" of cognitive processes. In this paper, we show that even fully distributed systems -i.e. systems with no modularity can give rise to double dissociations. We give the example of a recurrent neural network which draws loops and spirals which shows a double dissociation between the two tasks when lesioned. This result suggests that the observation of a double dissociation implies little about the modularity of the underlying system. In the final section we argue that a dual task technique can give additional hints about the structure of the underlying system because the class of distributed systems we describe are not able, in general, to perform two tasks at the same time. Finally, we argue that neurobiology has to be taken into account in order to interpret purely behavioral data.

Introduction

For several centuries neurological patients has been used to inform and constrain psychological accounts of normal function. Such evidence has served both to test existing psychological theories and to suggest how new theories can be developed. The value of neuropsychological evidence in theory testing is at least relatively uncontroversial and has been carefully analyzed (Caramazza, 1986). Much of the current upsurge of interest in neuropsychology within cognitive psychology and cognitive science has, however, stemmed from the hope that studying impaired function can play a role in the construction of theories of normal function. For example, a central theme of Shallice's recent and important book (Shallice, 1988) is that cognitive neuropsychology can have an important and proactive input into building theories of the processes involved across the range of cognitive domains, from language, memory and thinking to perception and action. We share the hope that evidence from impaired function may be an important and much needed source of constraint on cognitive theory, but suspect that inferences from neuropsychological data should be used to guide the development of new theories only with considerable trepidation Other authors (Henderson 1981; Crowder 1982), express of similar sentiments, though for different reasons). In the context of theory testing, Caramazza (Caramazza 1986) has argued that dissociations and associations are equally important. In the context of theory construction, Shallice (Shallice 1988) has argued that dissociations, and, in particular, double dissociation are particularly valuable, since they may reveal the gross structure or "modularity" of cognitive processes. In this paper, we argue that inference from double dissociation to a particular modular structure of the underlying cognitive system is problematic because double dissociations can be observed in a fully distributed system - that is, a system which does not decompose into isolable subsystems.

Even if double dissociations *per se* are not, as we argue, a sure guide to the existence of separable underlying subsystems, it may be that for example, a dissociation between "phonological" and "lexical" reading strategies do indicate the existence of distinct phonological and lexical routes. The plausibility of the dual route versus a single route model can only be decided in the light of the relative merits of specific models attempting to account for the range of data from impaired and normal function (for example, Patterson, Seidenberg & McClelland1989; Coltheart 1990). one piece of data for which reading models must account will be the dissociation of phonological and lexical reading. Our contention is not that double dissociations do not amount to interesting data for theory development and evaluation - rather we argue that they have no special status as a means of directly uncovering the modularity of the cognitive system.

Single and double dissociations

The range of characterizations of the method of double dissociation (Teuber 1955; Kinsbourne 1971; Shallice 1988) makes exposition of the method difficult. We shall assume what we take to be a typical modern "functional" formulation.

A patient with a lesion exhibits a single dissociation between tasks I and II when performance on task I is very poor, whereas performance on task II is either close to or at a normal level, or at least very much better than performance in task I (Marin, Saffran & Schwarz 1976; Beauvois & Derouesne 1979; Shallice 1988). It was once thought that such dissociations allowed one to infer that the set of isolable processes underlying the two tasks must be different. However, it has been argued that this inference is not licensed, since task I may make greater demands on a single damaged subsystem(s) than does task II. A subsystem working at, say, 50% capacity might be adequate for task II, but not sufficient for task I. This is often referred to as the problem of resource artefacts (Shallice 1988; Dunn & Kirsner 1988). In response to such

difficulties, it has been proposed that double rather than single dissociations are required to infer that two tasks draw on different processes, subsystems or modules. Tasks I and II doubly dissociate if there are patients A and B, such that A is more impaired than B in task I and, conversely, B is more impaired than A in task II. The point is that, unlike single dissociation, double dissociation cannot be generated with a resource artefact explanation. If task I makes greater demands on a single processing subsystem than task II, then task II may be selectively preserved (generating a single dissociation), but the reverse cannot occur. For if the subsystem is sufficiently impaired to damage task II, then task I, which relies on it even more heavily, will be even more severely impaired.

Original formulations of the inference from double dissociation (Teuber 1955; Kinsbourne 1971) assumed distinct and consistent lesion sites for patients with each kind of selective impairment. This anatomical assumption has been dropped in more recent "functional" formulations (Marin Saffran & Schwarz 1976; Shallice 1979; Shallice 1988). We argue that this less stringent criterion, although widely used (Caramazza 1990), may suggest an entirely misleading picture of the modularity of the underlying system.

Double dissociations in distributed systems

Shallice (Shallice 1988) observes that "to make the inference [from observed double dissociation to separate underlying subsystems] valid, one would need to add the assumption that [double] dissociations do not arise from damage to non-modular systems" (Shallice, 1988:248). *Prima facie*, this claim runs counter to evidence for double dissociation in lesioned distributed neural networks (Wood 1978; Wood 1980; Wood 1982; Gordon 1982; these studies are based on the "brain state in a box" model of Anderson, Silverstein, Ritz & Jones 1977).

Wood specifies two patterns for the network to learn, which differ in activation at just two of the units. Selective ablation of each unit produces selective damage to the memory for each pattern. Thus the memory for the two patterns doubly dissociates, even though the memory for each pattern is distributed through entire set of network weights. Such examples rely on a close relationship between the structure of the task and particular units. Shallice argues that this may reduce their relevance to discussion of effects of damage on real neural networks which "...will be composed of millions of neurons... [of which] no individual neuron is likely to have much importance in determining what output occurs" (Shallice 1988:255). He thus adds two conditions that a lesion in a distributed system would have to meet to be threatening to the double dissociation inference: "First, before the lesion is made, the influence of any particular neuron on what output is produced should be small. Second, the neurons affected by the lesion should not be selected by some complex algorithm that is determined by the dissociation to be explained and that is not typical of those that arise naturally. It seems most unlikely that if these conditions are satisfied, a ...double dissociation could be demonstrated in a properly distributed memory system." (Shallice 1988:256).

While this claim is couched in terms of memory, the domain in which Wood's network showed a double dissociation, it is clearly intended to apply more generally. We now describe a fully distributed system which performs two tasks - drawing loops or spirals, and which is intended to satisfy Shallice's two conditions. We shall introduce the network in three stages. Firstly, we describe the non-linear equation which the network implements. Secondly, we give a simple "local" network implementation of the equation. Finally, a totally distributed version of this network is described, and the effect of damage to this network discussed.

The equation

The equation that we chose is a simple iterative difference equation, a variant of the logistic difference equation, $x(n+1) = \text{lambda}.x(n).(1-x(n))$, used in classical population genetics (e.g., Maynard-Smith 1968) which was unexpectedly found to produce chaotic behavior (May 1976). Our equation is the delayed logistic map: $x(n+1) = \text{lambda}.x(n).(1-x(n-1))$, which, when $x(n+1)$ is plotted against $x(n)$ for each n produces either spirals if lambda is less than 2 and loops if lambda is greater than 2 (fig 1).

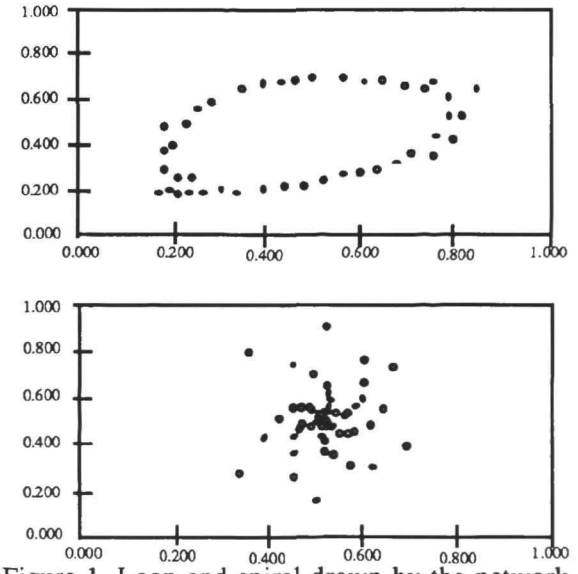

Figure 1. Loop and spiral drawn by the network (lambda equal to 1.925 and 2.075 respectively).

This critical value, at which the topology of the output is, in dynamical systems terminology, "structurally unstable", is know as the Niemark bifurcation. This equation is a standard example in non-linear dynamics (for example, Thompson & Stewart 1986). The parameter lambda, which determines whether a loop or a spiral is drawn, and the particular form of each, is what we shall term the "global" feedback parameter of our distributed network, and

perturbation of this parameter due to damage can lead to the selective loss of either spiral or loop drawing.

A local network implementation

A natural implementation of this equation in a simple neural network is shown in Figure 2.

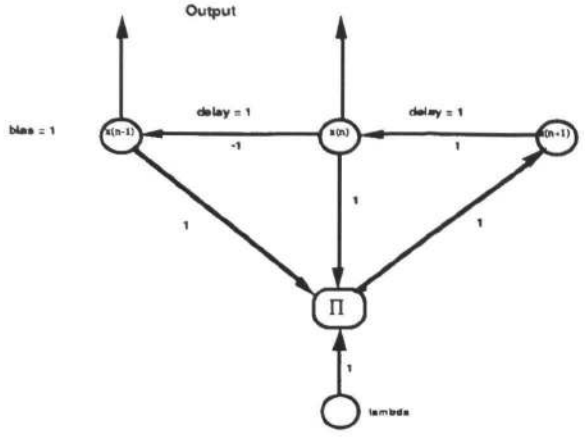

Figure 2. Local implementation of the delayed logistic function.

To draw a spiral a lambda of less than 2, say 1.95, is necessary. A lambda of 2.05 might be used to draw a loop.

The structure of the network mirrors the fact that the equation embodies both non-linearity and feedback. All units are linear, except for the single multiplicative unit (Hinton 1981), which takes the product of the value of the input, and the two previous values x(n-1) and x(n). The delay-lines feed the activation of the x(n) unit to the x(n-1) unit at the next time step, and the activation of the x(n+1) back to the x(n) unit. All other lines propagate immediately (with delay 0). The delay lines serve to "feed back" the output of the units at a given time-step as inputs at the next time-step. The combination of non-linearity and feedback is required for a system to exhibit a range of dynamically interesting behaviors, including chaotic behavior. It is not important here that we are using a discrete iterative map, rather than a system which continuously evolves in time according to a set of differential equations. The advantage of iterative maps, exploited both in the study of non-linear dynamics and in our illustrative example, is simplicity.

A distributed network implementation

In the local implementation, the loss of any unit or link would lead to catastrophic failure. There are, of course, many ways in which this function could be distributed across a larger number of units - we choose one of the simplest. For each local element, a corresponding distributed network will have the following properties (Chater & Ganis 1991) :

i) Each single unit of the local network is replaced by a group of units in the distributed network. The mean activation value of all the units in a group corresponding to the activation of the single unit in the local implementation.

ii) Each single link between a pair of units is replaced family of links, such that there is a connection between each unit in the "source" group to each unit in the "target" group. Thus, each unit in the target group is not fed a substantial input from a single unit in the input group, but rather receives a small amount of input from every unit in the source group (fig3).

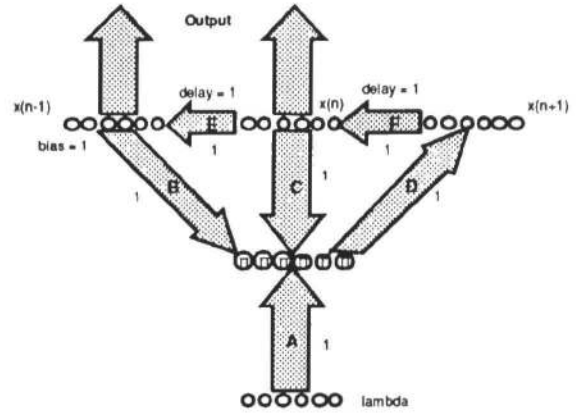

Figure 3. Distributed implementation of the delayed logistic function.

An average (Chater & Ganis 1991) of the values of these links corresponds to the value on the corresponding single link of the local implementation (there is the additional constraint that the values of the links from each unit are the same, to keep analysis tractable).

Thus, the distributed implementation (figure 2) simply reduplicates hardware so that the role of each unit is replaced by a set of units. The values of x(n-1), x(n) and x(n+1) are represented by the average of the activation levels of each group of unit, and the large arrows denote totally connected sets of links between groups of units with the ensemble averages (1 or -1) as shown.

By design, the distributed network behaves in the same way as the corresponding local network the ensemble average values of groups of units and links correspond to the values of single units and links in the local implementation. If we consider the various ways in which the network may be impaired by ablation of a subset of a single group of links or set of units, then the damaged network will again be equivalent to a local network. Suppose that some proportion of a set of links with ensemble average 1 are ablated. If the links damaged are mostly positive, then the the ensemble average of the set will be reduced - the equivalent local network will have a single link with a value of, say, 0.8. On the other hand, if it happens that the links ablated are mostly negative then the ensemble average of the set (and hence the value on the corresponding link in the local implementation) may be increased - say to 1.2. The way in which positive (excitatory) and negative (inhibitory) connections are anatomically organised (whether, for example, they are separated or together), and depending on the typical size of the link values (e.g., whether the inputs to a typical unit are, say, 0.2, -0.2, 0.4, -0.1, 0.4, -0.2, 0.3,

0.2 (= 1) or 1, -5, 6, 2, 3, -3, 4, -7 (=1)) will have a dramatic impact on the distribution of changes in ensemble average that a lesion is likely to cause. We shall consider a full range of lesions at each of the links A, B, C, D, E & F and ablations of units in each set of the units. Notice that the ablation of a unit is equivalent to the removal of the connections which feed out of that unit. Thus unit ablation is simply a special case of the ablation of connections A-F, except for the ablation of the x(n) group, which feeds into both x(n-1) and the multiplicative units thus the ablation of these units is equivalent to the ablation of subsets of two sets of links - C & E. Suppose that the value of a link is reduced/increased by a factor of mu (mu1 and mu2 in the C & E case). Then the equations governing the impaired system is no longer $x(n+1) = lambda.x(n).(1-x(n-1))$ but:

A, B, C, D, F: $x(n+1) = [lambda*x(n)*(1-x(n-1)]*mu$

E: $x(n+1) = lambda * x(n) * (1-mu) * x(n-1)$

C & E: $x(n+1) = [lambda*x(n)*(1-mu2*x(n-1)]*mu$

Thus the general form of the equations after damage is:

$x(n+1) = [alpha * lambda] * x(n) * (1-beta * x(n-1)$

The beta term, which differs from 1 only in two cases, serves only to slightly distort the spirals or loops drawn, rather than changing their underlying structure. In particular, it does not disturb the value of the Niemark bifurcatiom. So, if the feedback parameter lambda is less than 2, a spiral will be drawn; if lambda is greater than 2, a loop will be produced. Hence, damage to the distributed network has the same effect as setting lambda to lambda * mu, in the corresponding local network. Now it is clear how a double dissociation may arise.

Suppose spirals are drawn with lambda 1.95, and loops with lambda 2.05. Damage to the network which ablates more excitatory than inhibitory connections will reduce the amount of feedback, and mean that alpha is less than 1 - say 0.95. In this case, a lambda of 2.05 would correspond to lambda*0.95, that is 1.95 : the network will no longer be able to draw loops.

Notice that our example conforms with Shallice's strictures. Firstly, the influence of each neuron on task performance is small. Whereas in Wood's examples, particular neurons were especially significant for remembering certain patterns, in this example each neuron has the same influence every other neuron, in both loop and spiral drawing. Damage produces a dissociation by changing global system parameters rather selectively impairing particularly important individual units. Secondly, the kinds of damage that we have suggested do not involve any complex procedure for selecting which parts of the net should be ablated. Any kind of damage which alters the amount of feedback in the system, whether as a result of a chemical changes, loss of some external non-specific input or some other pathology, is liable to generate a dissociation of loop and spiral drawing.

Conclusion

We have shown in this paper that fully distributed systems can generate double dissociations. The fact that one distributed system can produce a double dissociation does not necessarily mean that a large and neurally plausible class of distributed systems can do so, and only in the latter case will the inference from double dissociation to modularity be impugned in practice. Certainly neural systems do fall into the class of non-linear dynamical systems with feedback, and will exhibit far more interesting and elaborate dynamics than that generated by the logistic function. In particular, they may have a far more elaborate range of distinct structural configurations rather than just two, and the operative configuration is likely to be determined by a complex set of global parameters rather than the value of a single parameter, as here. Further, which of these structural configurations is operative may well importantly affect the task performed, and the relevant parameter values may be altered by a range of neurologically plausible forms of damage - chemical imbalances, localized lesions, loss of non-specific input from other centers, and so on. There is strong evidence that real neural networks can be multifunctional; this means that a single anatomically defined neural network can generate more than one behavior , depending on the value of one or more global parameters. 'Modulation of the network, synaptic, and cellular building blocks can serve to adapt the output pattern to ongoing needs or may dramatically reorganize a network into an entirely new mode mediating a different behavior' (Getting 1989). Therefore, there seems to be plenty of scope for double dissociations in real distributed neural systems. It may be countered that, whereas it is easy to see how loop drawing and spiral drawing may be products of the same system, with different global parameter values, it is less easy to conceive of how different global parameter values might transform a system from, say, phonological reading, to lexical reading. On the other hand, it is only easy to see how different parameter values of the same underlying system can produce loops and spirals in retrospect, and that our understanding of the properties of complex non-linear dynamical systems is too slight to put much weight on intuition. It may be that confidence in the inference from observed dissociation to underlying modular organization is based as much on our current lack of understanding of distributed systems, as on their underlying properties.

The question of how we can reduce the uncertainty of inferences from double dissociations obviously arises. We suggest two answers, one tied to our particular example, the other more general.

With regard to our specific example, we suggest that data from dual task techniques might be useful in interpreting double dissociations. Indeed, the class of multifunctional networks we have described is not able to perform two tasks at the same time (namely, drawing loops and spirals). Therefore, if there is a double dissociation between two tasks, and the subjects are able to perform both tasks at the same time, it seems unlikely that the underlying system be a multifunctional network like the one we put forward. In practice, results from dual task experiments may be difficult to interpret; for example, it is obvious that it is impossible to draw spirals and loops with the same hand at the same time; this, however, does not imply anything about the

existence of one or more cognitive isolable subsystems for drawing spirals and loops.

With regard to the general problem of interpreting double dissociations, we think that the only possible way to reduce uncertainty is to take into account evolutionary constraints and advantages (Weiskrantz 1990) and data from neuroscience (Weiskrantz 1990; Sereno 1990). Indeed, we think that the purely 'functional approach', still dominant in cognitive neuropsychology, has led to '...a kind of candy floss neuropsychology, brightly labelled, complexly reticulated, full of growth but shifting in substance' (Weiskrantz 1990), the reason for this being the systematic overlooking of neurobiological data'.

Acknowledgements

This work was carried out while the authors were at the Department of Psychology, University College London. We should like to thank Peter Passmore for discussion and help with maths packages, and thanks also to Margaret and Marty Sereno for comments on the manuscript.

References

Anderson, J. A., Silverstein, J. W., Ritz, S.A. & Jones, R. S. 1977 . 'Distinctive features, categorical perception, and probability learning: Some applications of a neural model. '.*Psychological Review* 84:413-451.

Beauvois, M. F. & Derouesne, J. 1979 'Phonological alexia: Three dissociations', *Journal of Neurology, Neurosurgery and Psychiatry* 42:1115-1124.

Chater, N.; Ganis, G. 1991. Can double dissociations uncover the modulairy of the cognitive system? Forthcoming.

Caramazza, A. 1986 . On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single patient studies. *Brain & Cognition* 5:41-66.

Coltheart, M. 1990. Computational models of the human reading system . Paper presented at the British Psychology Society: Cognitive Psychology Section Annual Conference, University of Leicester, September, 1990.

Crowder, R. G. 1982. General forgetting theory and the locus of amnesia . In L. S. Cermak (ed.), *Human memory and amnesia* Hillsdale, N.J.:Earlbaum.

Dunn, C.J. & Kirsner, K. 1988. Discovering functionally independent mental processes: the principle of reversed association. *Psychological Review*. 95:91-101.

Getting, P. A. 1989 Emerging principles governing the operation of neural networks. *Annual Review of Neuroscience* 12:185-204.

Gordon, B. 1982. Confrontion naming: Computational model and disconnection simulation In M. A. Arbib, D. Caplan and J. C. Marshall (eds.).*Neural models of language processes* New York : Academic Press.

Henderson, L. 1981. Information processing approaches to acquired dyslexia. *Quarterly Journal of Experimental Psychology* 35A:507- 522.

Hinton, G. E. 1981. A parallel computation that assigns canonical object-based frames of reference. In Proceedings of the 7th JointConference on Artificial Intelligence, 683-685.

Kinsbourne, M. 1971. Cognitive deficit: Experimental analysis. In J. L. McGaugh (ed.). *Psychobiology* New York, Academic Press.

Marin, O. S. M., Saffran, E. M. & Schwarz, D. F. 1976. ' Dissociation of language in aphasia: Implications for normal functions. *Annals of the New York Academy of Sciences*, 280:868-884.

May, R. 1976 . Simple mathematical models with very complicated dynamics . *Nature* 261:459-467.

Maynard Smith, J. 1986. *Mathematical ideas in biology* . Cambridge: CambridgeUniversity Press.

McClelland, J. L. 1986. A programmable blackboard model of reading. In J. L. McClelland and D. E. Rumelhart (eds.) *Parallel distributedprocessing: explorations in the microstructures of cognition. Volume 2: Psychological and biological models.* MIT Press, Cambridge, Mass..

Patterson, K. E., Seidenberg, M. S. & McClelland, J. L. 1989 . Connections and disconnections: acquired dyslexia in a computationalmodel of reading processes . In R. G. M. Morris (ed.) *Paralleldistributed processing: Implications for psychology and neurobiology.* Oxford:Oxford University Press.

Sereno, M.I. 1990. Language and the primate brain .In Proceedings of the 13th Annual Meeting of the Cognitive Science Society. Forthcoming.

Shallice, T. 1984 . More functionally isolable systems but fewer "modules" . *Cognition* 17:243-252.

Shallice, T. 1988. *From neuropsychology to mental structure* . Cambridge:Cambridge University Press.

Teuber, H.L. 1955. Physiological psychology. *Annual Review of Psychology.* 9:267-296.

Thompson, J. M. & Stewart, H. B. 1986. *Nonlinear dynamics and chaos* . Chichester, Wiley.

Weiskrantz, L. 1990. Multiple memory systems... *Proceedings of the Royal Society of London, Series B: Biological Sciences.* 1253:99-107.

Wood, C. C. 1978. Variations on a theme of Lashley: Lesion experiments on the neural network of Anderson, Silverstein, Ritz & Jones. *Psychological Review* 85:582-591.

Wood, C. C. 1980. Interpretation of real and simulated lesion experiments . *Psychological Review* 87:474-476.

Wood, C. C. 1982. Implications of simulated lesion experiments for the interpretation of lesions in real nervous systems . In M. A. Arbib, D. Caplan and J. C. Marshall (eds.), *Neural models of language processes.* New York:Academic Press.