

Double dissociation and isolable cognitive processes

Nick Chater

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

Giorgio Ganis

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

Abstract

Data from Neuropsychology have been widely used in order both to test pre-existing cognitive theories and to develop new accounts. Indeed, several theorists have used dissociations, and in particular double dissociations, both in theory testing and in developing new theoretical accounts

Double dissociations are indeed believed to be a key tool in revealing the gross structure or "modularity" of cognitive processes. In this paper, in the light of a case study in which a simple electrical system is systematically lesioned, we argue that double dissociation in an arbitrary modular system need not, and typically will not, reveal that modularity. These results suggest that the observation of a double dissociation implies little about the structure of the underlying system. We finish arguing that the weakness of the methods described involves that neurobiological data have to be seriously taken into account in order to uncover the real structure of the cognitive system.

Introduction

Evidence from brain damaged patients has been widely used in order to inform and constrain theories of normal functioning. Such evidence has served both to test existing psychological theories and to suggest how new theories can be developed.

The major aim of cognitive neuropsychology is to construct theories of normal function by studying impaired function (Caramazza 1986; Shallice 1988). The main assumption here is that patterns of impaired performance are not random but are constrained by the actual structure of the cognitive system itself. This is a typical inverse problem which, by definition, allows more than a single solution. The strategy followed by cognitive neuropsychologists in order to solve this problem has been to look for particular patterns of impaired performance, like associations and dissociations (Caramazza 1986; Shallice 1988). Double dissociations, in particular, are considered to be extremely valuable, since they may reveal the gross structure, or 'modularity', of the cognitive system

In a previous paper (Ganis & Chater 1991), we argued that inference from double dissociation to a particular modular structure of the underlying system is uncertain because double dissociations may be observed in a fully distributed system -that is, a system which does not decompose into isolable subsystems.

In this paper, we show that even in a modular system, double dissociation data, as conventionally interpreted, may be entirely misleading about that modularity.

Single and double dissociations

The range of characterizations of the method of double dissociation (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 (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). An analogy with locomotion may be helpful. There is a frequently observed dissociation between running and walking -many people who can walk normally can run hardly at all. This is not because running and walking use different bodily structures, but simply that running places greater resource demands upon those structures. Hence people with limited air intake, stamina and so on are able to walk much better than they can run. 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 subsystems. 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. For example, a double dissociation between running and throwing is found between patients with broken limbs. Those with broken legs have impaired running, but throwing is preserved, and those with broken (dominant) arm have selectively impaired throwing. The double dissociation inference is that there must be separate systems underlying throwing and running. *Prima facie*, these correspond to the sites of the damage, the arms and legs.

If double dissociation is to constitute an advance on single dissociation as a method for uncovering isolable subsystems, then it should not be obtained between running and walking, which are subserved by the same processing systems. Indeed, while there are patients who can walk but not run, there are presumably no patients who can run but not walk. Thus, unlike single dissociation, double dissociation cannot be generated with a resource artifact 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.

The method of double dissociation has pleasingly distinguished between the two examples, one in which the two tasks are subserved by different processes, subsystems or modules and one in which they are subserved by the same system. However, we shall argue that this apparently clear-cut outcome may be the exception rather than the rule.

Can double dissociations suggest the wrong modularity ?

In a previous paper (Ganis & Chater 1991) we have observed that the double dissociation inference may be misleading, since at least some double dissociations are consistent with a non-modular underlying architecture. We now argue that the inference can be misleading for modular systems too - the modularity postulated by the inference from double dissociation may have no relation to the actual modularity of the damaged system. In particular, we show that even under the weaker definition of modularity e.g. functional modularity (Shallice, 1988) the double dissociation method can be seriously misleading.

Let us consider a non-psychological example. Consider a dissociation between javelin throwing and rope climbing, due to blisters on the hands. Certain blisters (perhaps caused by overzealous javelin throwing on the previous day) will make throwing impossible, while leaving rope climbing more or less unimpaired. On the other hand, blisters in slightly different locations (perhaps caused by excessive recent rope climbing) will make climbing impossible will leaving throwing relatively intact. The double dissociation inference is that there must be separate processes or modules underlying rope climbing and javelin throwing. However, both involve the same hand-arm system (and indeed both involve the cooperative action of the whole of the body). It may be relatively easy to avoid being misled, if the source of damage is visible and its effects relatively easy to discern, as in the case of blisters and broken limbs. It may be very much harder when the damage, and the functional impact that the damage has, is only indirectly detectable. Of course, this is no more than a suggestive analogy.

To better assess whether the analogy is instructive or misleading, let us consider the specific example of a simple mechanism which performs two tasks. Task I is to repeat on-line each member of a sequence of "a"s and "b"s which are given as input. Task II is to rehearse whichever is presented at a specific time, and to ignore subsequent input. In its normal state, the system under study is able to switch between tasks I and II at will by varying the value of a parameter. After subjecting the system to a full range of possible lesions, the following pattern of performance was obtained (Table I):

		Task I (Repetition)	
		normal	impaired
Task II (Rehearsal)	normal	normal behaviour	5
	impaired	9	16

Table I

The table shows a clear double dissociation between tasks I and II. 5 lesions resulted in the selective loss of the ability to perform Task I, and 9 lesions resulted in the selective loss of the ability to perform Task II. 16 lesions result in the loss of performance on both tasks. As a first attempt to explain this data, the inference from double dissociation suggests that we can posit two distinct modules, one responsible for repetition and one for rehearsal. We might, for example, follow the logic which lead Marshall & Newcombe (1973) to postulate a "two-route" model of the reading process, giving a diagram such as (Figure 1).

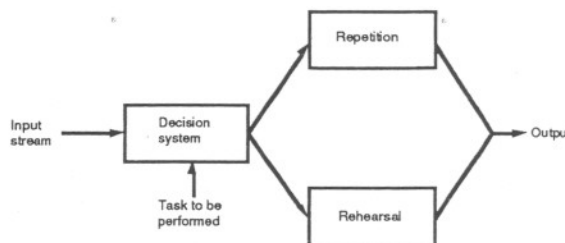


Figure 1 A boxes model inferred from the patterns of impaired performance in Table 1.

Selective damage to either the repetition or rehearsal modules would lead to the double dissociation between performance in tasks I and II. Damage to the decision system, which selects one of the other two modules as appropriate, might be the most likely explanation for the loss of both capabilities. This model may be assessed more closely by looking in more detail at the kind of pathologies found.

[1] Task I lost; Task II intact (5) No repetition (5): Task II intact; Task I locks to b after the first b in the input stream.

[2] Task II lost; Task I intact (9) No rehearsal (4): Task I intact; continuous stream of "a"s (4) or "b"s (5) in Task II.

[3] Both Task I and Task II lost (16)

[3a] Fixation (11): system may produce either only "a"s (10) or only "b"s (1), for both tasks.

[3b] Fixation and oscillation (1). Continuous stream of "a"s in Task I. Continuous stream of "...abababababa..." in Task II.

[3c] Mirroring and oscillation (1). In task 1, the output is a "mirror image" of the input. For example, "aabab" becomes "bbaba". Continuous stream of "...abababababa..." in Task II.

Let us consider each of [1]-[3]. in turn, with respect to our hypothetical model.

The basic pattern of [1] is explained by selective damage to the repetition module. That the damaged module produces only a continuous stream of output in Task I (b's) may also suggest various hypotheses about the possible internal organization of the module. If "a"s and "b"s are produced by different components of the module, then selective impairment of either, or a loss of the ability to switch between them, or some problem concerning the levels of their mutual inhibition might be responsible for this pathology. The fact that streams of "a" are not found may perhaps be a consequence of some asymmetry in this mechanism.

[2] is taken to be a consequence of selective damage to the rehearsal module. Again, that the damaged module produces only a continuous stream of output suggests that there may be separate components or submodules which produce "a"s and "b"s. As before, selective damage to either component, a loss of the ability to switch between them, or some problem concerning the levels of their mutual inhibition might be responsible for fixation on a single output. This suggests that, in key respects, the rehearsal and repetition modules may have a similar structure.

[3] Loss of the ability to perform both tasks may be due either to damage to both the repetition module and the rehearsal modules, or to the decision box. [3a] A continuous fixation on the same letter for both tasks, a very common pathology observed with nine distinct lesions, might result from damage to the decision box disconnecting the modules from current input, and any instructions to change between tasks. The current output cannot be disabled, and may hence be repeated indefinitely. There are just two lesions with puzzling behavior for this account. [3b] and [3c] produce continuous oscillations between "a" and "b", on task II. [3c] also "mirrors" rather than repeats its input on task I. It may be hoped that a fuller understanding of the fine details of each module and its component parts might reveal the origins of such pathologies, but that the basic structure of the hypothetical model is not impugned.

So, in the light of this reasoning we have derived a model with the following gross structure (fig 2):

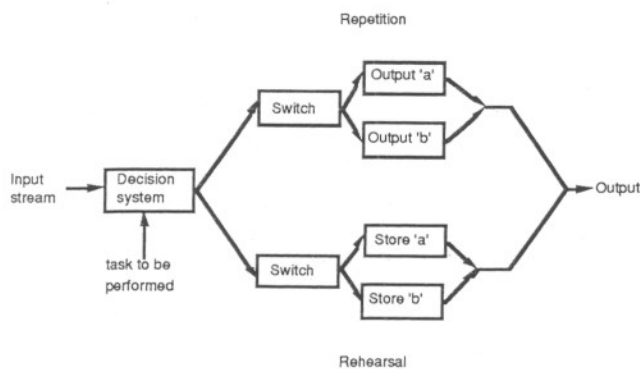


Figure 2. Detailed model derived from the the patterns of impaired performance in Table I.

A possible alternative is shown in figure 3 and there are, of course, many others.

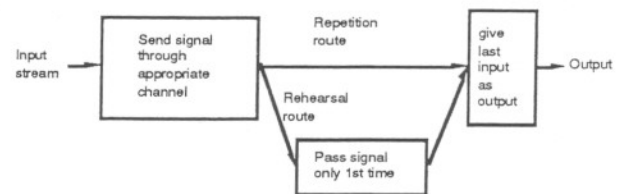


Figure 3. Another possible model.

If the task is repetition, the input stream is directed either directly to the output module. To perform rehearsal, input is directed to the the "block" module, which passes only its first input and blocks all future inputs. The output module produces an "a" if its most recent input was an "a", and a "b" if its most recent was a "b". So, if the system is in repetition mode, the output module receives a continuous stream of input and simply reproduces that stream. When the system is in rehearsal mode, only a single input is received, from the block module, and the output module repeats that input endlessly. The particular models that we have suggested are of illustrative significance only. In each, the basic finding of double dissociation and certain other aspects of the data may be accommodated, with suitable auxiliary assumptions. Certain aspects of the data are left unexplained by all (for example, oscillations in output, and the production of the "mirror" of the input stream). What is significant about the models that we have discussed, and variants like them, also derived to explain the double dissociation is that they bear no relation to the modularity of the mechanism which generated the data: a 'pseudo D' flip-flop. Figure 4 shows the circuit diagram of such a 'pseudo D' flip-flop, the basic component of memory registers in digital computers, which is able to hold a single bit of information.

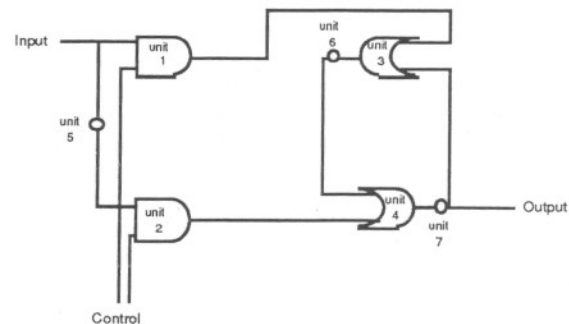


Figure 4. Diagram of a 'pseudo D' flip-flop.

The input and output streams of "a"s and "b"s is, in this context, usually thought of as streams of "0"s and "1"s. When the control signal is on (the device is in "repetition" mode) the input is passed unchanged to the output. When the control signal is off (the device is in "rehearsal" mode), the input is cut off, and the output is frozen to its previous value. The circuit can be interpreted either in a functional or in an anatomical way

(i.e., each unit might not correspond to any particular piece of hardware; alternatively, it could correspond to a silicon or neural gate).

This system is highly modular and has 7 component units or gates, connected by a number of links. Values for links and units may be idealized as having only two discrete values, 0 and 1. The two AND gates on the left hand side of the diagram, output a 1 just when both their inputs are 1 and a 0 otherwise. The two OR gates on the right output a 1 if at least one of their inputs is a 1, and 0 otherwise. There are also three NOT gates (denoted by small unfilled circles) which take a single input, mapping 0 onto 1 and 1 onto 0.

Prima facie, the number, function and interconnection of the components of the actual system bear little resemblance to the models proposed above, in an attempt to account for the data on pathological function. Crucially, there appear to be no separate subsystems which subserve the two tasks of rehearsal and repetition, and the double dissociation inference - the system is composed of four components, rather than the two or three postulated above, and each is a simple logic gate rather than having more elaborate functions such as gating the input or printing out the last item received.

It might be thought that the circuit diagram representation is simply too detailed to reveal a clear modularity. Perhaps there is some subsystem of units and connections particular to each task.

Consideration of how the flip-flop operates does indeed suggest, for example, that the right hand side of the diagram - the OR units 3 and 4 - is concerned with rehearsal (since it embodies a loop in which activation values can be stored), and the left hand side - the AND units 1 and 2 - is concerned with passing on, or gating out, the incoming input, depending on whether the task is repetition or rehearsal (input is only passed on if the control signal is 1). This gives us Figure 5, which, hearteningly, is not so very different from Figure 3.

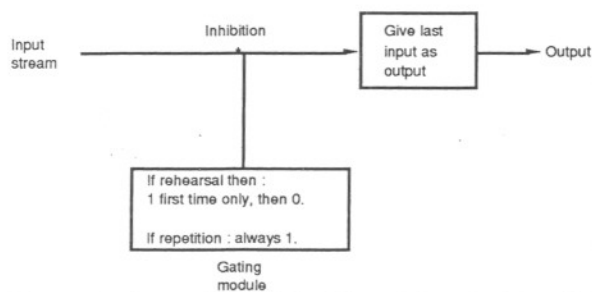


Figure 5 A possible 'higher' level description of the flip-flop in Fig 3.

However, Figure 5 is precisely the kind of diagram that appears to be incompatible with double dissociation evidence (still less can it explain the detailed patterns of pathologies observed - oscillations, mirroring the input stream, and so on). Damage to module A, cutting off or distorting the input stream, seems likely to impair both repetition and rehearsal (since the single input which must be stored may not be passed successfully). Damage to

module B, which is also implicated in both tasks, should impair the ability to printing of the most recent input, whether in repetition or rehearsal. In short, to the extent that there is a higher level modular structure, each part of that structure is involved in both tasks, pace the double dissociation evidence. So, not only can the double dissociation inference lead to the postulation of high level descriptions unrelated to the real components of the system, but it may specifically preclude the high level descriptions which are closest to the truth.

Conclusion

We have shown in this paper that double dissociations can suggest the wrong modularity even if the underlying system is modular.

It is worth discussing why the double dissociation method does not work with our simplified examples. The double dissociation method is based on the assumption that the cognitive system is composed of a set of modules the effectiveness of operation of which can be measured by a single real variable - the resource. The 'neurological disease depletes the amount of each particular resource by an unknown value, which varies with the patient' (Shallice 1988). The performance on a given task, say delayed repetition, is then assumed to be a function of these real-valued numbers, one for each module. Even with a simple example like the flip-flop one, this approach has severe problems. Take, for example, an AND gate. There are different ways in which it can be damaged (Chater & Ganis 1991). Each kind of damage can produce a particular 'deficit' in the AND. Theoretically, the AND can become one of the 16 possible two-arguments binary gates (the extreme cases being a gate which gives as output always 1 or always 0). The impact of these different damages on the performance of the flip-flop in the two tasks is dramatically different even when the severity of the damage (in terms of percentage of spared input-outputs mappings) is kept constant. This is why a resource/performance analysis cannot account for the behavior of the flip-flop. More realistic systems seem to be even more sensitive to the details of the damage (Getting, 1989). We think that the examples so far discussed are just instances of the difficulties which top-down functional investigations of the cognitive system have run into. In fact, the claim that a pure functional approach has to work in parallel with neurobiological investigation is often made (see for example Caramazza(1990)). Unfortunately, these claims are not followed by concrete efforts in trying to base neuropsychological theories on neurobiological data. A possible reason is that, as in our examples, the wrong modularity has been inferred, rendering simply impossible any link with neurobiological data (see also Seidenberg (1988) and Sereno (1991)).

Appendix

The sites of the lesions are shown in figure 6.

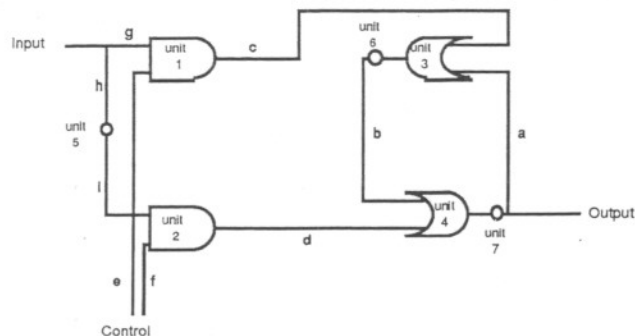


Figure 6. Sites of the 'lesions'.

A summary of the pathologies obtained with the various lesions is shown in table II

Pattern of performance	Details	Lesion responsible
Selective loss of task I (no repetition)	In task I, output locks to I after first 'I' in input	2.0 4.i d f i
Selective loss of task II (no rehearsal)	In task II, output has a fixed value	Fixation on '0': a 3a 3.i 2.i Fixation on '1': b 1.1 3.1 4.ii 1.i
Loss of tasks I and II	The output has a fixed value, either '0' or '1' in both tasks.	Fixation on '0': c e g h 1.0 2.1 2.ii 3.0 3.ii 4.1 Fixation on '1': 4.0
	Fixation and oscillation: Fixation on a single value in one task; instability and oscillation in the other	1.ii 2.ii n2 n3
	Mirror and oscillation: in task I the output is the negation of the input. Oscillation in task ii.	n4

Table II

We have attempted to pay enough regard to the details of electrical circuitry. The lesions correspond, at least roughly, to the kinds of damage that can occur in a real circuit. However, since illustrative value rather than electronic realism is of our goal, certain simplifications have been introduced. The behavior of the circuit is considered with each of the wires cut in turn. We make the simplifying assumption that cutting a wire corresponds to a zero input into the component into which it feeds. Damage to both AND and OR gates gives rise to the same four distinct pathologies (although for rather different reasons). Output may be fixed to 0 or 1 or may simply pass one of the inputs i or ii. Such damage, for unit X, is denoted X.0, X.1, X.i and X.ii respectively. For simplicity, only one form of damage to NOT gates is considered - a short-circuit from input to output wire. The damaged NOT gate thus passes the input unchanged. Ablation of the unit, so that its output is fixed at 0 produces no new patterns of behavior of the circuit and is so not considered. Table II summarizes the input/output functions of the damaged gates.

i	ii	X0	X1	Xi	Xii
0	0	0	1	0	0
0	1	0	1	0	1
1	0	0	1	1	0
1	1	0	1	1	1

Table II.

The details of the various kinds of damage are described elsewhere (Chater & Ganis 1991).

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

- 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.
- Chater, N.; Ganis, G. 1991. What double dissociation tell us about the modularity of the cognitive system. Submitted to *Psychological Review*.
- Ganis, G.; Chater, N. 1991. Can double dissociation uncover the modularity of the cognitive system?. In Proceedings of the 13th Annual Meeting of the Cognitive Science Society. Forthcoming.
- Getting, P. A. 1989. Emerging principles governing the operation of neural networks. *Annual Review of Neuroscience* 12:185-204.
- 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.
- Marshall, J.C.; Newcombe, F. 1973. Patterns of paralexia: a psycholinguistics approach. *Journal of Psycholinguistics Research*, 2:175-199.
- Seidenberg, M.S 1988. Neuropsychology of language. Paper presented at the Venice Conference on Connectionism and Neuropsychology.
- 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.