Stochastic approaches to understanding dissociations in inflectional morphology

Kim Plunkett *, Stephan Bandelow

Department of Experimental Psychology, Oxford University, South Parks Road, Oxford OX1 3UD, UK

Accepted 22 April 2006
Available online 5 June 2006

Abstract

Computer modelling research has undermined the view that double dissociations in behaviour are sufficient to infer separability in the cognitive mechanisms underlying those behaviours. However, all these models employ multi-modal representational schemes, where functional specialisation of processing emerges from the training process. Targeted lesioning of different regions of functional specialisation leads to varied but predictable deficits in model performance. We argue that multi-modal representational schemes are not a necessary condition for the observation of double dissociations in an information processing system that shares resources across multiple tasks. Using a uni-modal representational system, we demonstrate that double dissociations may also result from stochastic processes. Lesioning experiments on a single-route, uni-modal connectionist model of regular and irregular noun and verb morphology confirm and extend earlier work demonstrating that selective impairment across tasks can result from damage to a distributed information processing system. A systematic investigation of the degree to which performance deteriorates across different inflectional classes reveals that simple and double dissociations can occur in this single-route, uni-modal model. An important prediction of the model is that double dissociations resulting from stochastic processes should be extremely rare. However, they are particularly likely to occur when the researcher uses test batteries consisting of a small number of items. Given that cognitive neuropsychologists rarely provide details about the distribution of performance in a disordered population, it is concluded that a stochastic interpretation of double dissociations may have wider applicability than is normally supposed.

Keywords: Double dissociation; Inflectional morphology; Past tense; English plurals; Connectionist models; Neural networks; Aphasia; Alzheimer’s disease; Parkinson’s disease; Single case studies

1. Introduction

Dissociations, and specifically double dissociations, are widely considered to be among the more powerful tools in a cognitive neuropsychologist’s arsenal. A double dissociation occurs when “...on task I, patient A performs significantly better than patient B, but on task II, the situation is reversed”. (Shallice, 1988, p. 235) Traditionally, a double dissociation has been interpreted as reliable evidence for identifying the separability of the mechanisms underlying two behaviours. Coltheart (1985) has referred to this approach as the “logic of cognitive neuropsychology”.

This type of reasoning has been applied in the domain of inflectional morphology. Double dissociations between regularly and irregularly inflected verbs have been reported in aphasic patients. Marslen-Wilson and Tyler (1997) demonstrated differential patterns of loss in priming of verb stems by regular and irregular past tense forms and concluded that the observed double dissociation indicated both a neurological dissociation and a differentiation in the type of mental computation for these two morphological categories. Likewise, Ullman et al. (1997) reported an investigation of anterior and posterior lesion aphasics, as well as patients suffering from either Alzheimer’s or Parkinson’s disease. In a past tense elicitation task, the anterior lesion...
aphasic and patients suffering from Parkinson’s disease had greater difficulty generating past tenses for non-words than irregular verbs. In contrast, the posterior lesion aphasics and patients suffering from Alzheimer’s disease had greater difficulty generating past tenses for irregular verbs than non-words. Ullman et al. (1997) argue that the double dissociation observed across this group of patients serves as evidence for a dissociation in the type and localisation of the mechanisms underlying regular and irregular inflection. Similar arguments have been made about differences in the reading of nonce words and real words (Coltheart, Curtis, Atkins, & Haller, 1993) or between the processing of abstract and concrete nouns (Warrington, 1981), to cite just two other well-known examples of double dissociations.

The “logic of cognitive neuropsychology” (Coltheart, 1985) has been challenged over the past decade by the construction of computational models that can demonstrate double dissociations in behavioural tasks, despite the fact that the modelled tasks share information processing resources (Farah & McClelland, 1991; Joanisse & Seidenberg, 1999; Plaut & Shallice, 1993). The demonstration that double dissociations can occur in systems without modularity of mechanism has prompted a reassessment of the theoretical underpinnings of cognitive neuropsychology (Plaut, 1995). In particular, it is claimed that the observation of a double dissociation is not sufficient grounds to claim modularity of mechanism, only the weaker claim of functional specialisation (Shallice, 1988).

Plaut and Shallice (1993) simulated a double dissociation between concrete word and abstract word reading reported in Warrington (1981) by lesioning physically distinct parts of a connectionist model of reading. Damage to connections emanating from the orthographic input units yielded reliably better performance on abstract words compared to concrete words, whereas damage to connections responsible for semantic attractors yielded reliably better performance on concrete words. Plaut and Shallice (1993) explained the double dissociation in terms of the extent to which concrete and abstract words rely on the information encoded in the connections in different pathways: Concrete words engaged the semantic attractors more effectively than abstract words in normal processing and were more disrupted than abstract words when this attractor pathway was lesioned. In effect, the semantic attractors were functionally specialised for processing concrete words. When the direct pathway from the orthographic units was damaged, an intact semantic attractor pathway could compensate for the degraded processing of concrete words. No such compensation was afforded to abstract word reading.

Joanisse and Seidenberg (1999) simulated a double dissociation between regular and irregular verb morphology. They described a connectionist model that utilised phonological and semantic information when inflecting verbs in the past tense. They found that damage to semantic information in the network resulted in greater impairment of irregular verb inflection whilst damage to phonological information resulted in greater impairment of the network’s capacity to generate past tenses for novel words. Joanisse and Seidenberg (1999) point to the similarity in performance of their neural network model and the breakdown in performance that Ullman et al. (1997) observed in aphasic patients and patients suffering from Parkinson’s or Alzheimer’s disease.

In both of these examples, double dissociations were demonstrated in non-modal systems by lesioning regions of the network which became functionally specialised as a result of training. In the models described by Plaut and Shallice (1993) and Joanisse and Seidenberg (1999), the emergent functional specialisation was predictable from architectural and representational constraints (Elman et al., 1996) imposed by the modeller. In particular, both models exploited multi-modal representational systems where distinct layers of processing units were assigned the task of representing orthographic, semantic or phonological aspects of the task. Lesions to connections directly attached to orthographic, semantic or phonological units resulted in deficits in performance which reflected the degree to which the modelled tasks relied upon orthography, semantics or phonology for their proper operation.

The question arises as to whether the multi-modal representations exploited by these models were a necessary condition for the occurrence of double dissociations. In other words, is it possible to observe double dissociations in uni-modal models where only a single type of representational coding is used, such as the Rumelhart and McClelland (1986) model of past tense inflection, where only phonological codings are used? Uni-modal connectionist simulations offer the most extreme examples in the literature of non-modal approaches to information processing systems since, by definition, there can be no division of labour between tasks based on distinct representational spaces. The occurrence of double dissociations in uni-modal connectionist networks would therefore constitute an even stronger argument against the logic of cognitive neuropsychology (Coltheart, 1985) than the demonstration of double dissociations in multi-modal connectionist networks (Farah & McClelland, 1991; Joanisse & Seidenberg, 1999; Plaut & Shallice, 1993) where the distinct orthographic, semantic and phonological representational spaces (though shared across tasks) constitute a half-way house between completely modular and fully distributed processing systems.

2. Uni-modal models of inflectional morphology

The use of uni-modal models of inflectional morphology does not require a theoretical commitment to the view that inflectional processes are purely phonological. For example, semantic information is essential to disambiguate the inflectional process in many cases, e.g., ring → rang/wrung/ ringed, hang → hung/hanged, etc. However, individuals can produce the plural or past tense of novel words (Berko, 1958). Clearly, they do not need semantic information to inflect words, even though they may often use it. The
purpose of investigating uni-modal models is to examine the
degree to which phonological processing by itself can
contribute to double dissociations. The only way to assess
this is to eliminate semantics and any other representa-
tional resources contributing to inflectional morphology.

Marchman (1993) trained a uni-modal, single-route net-
work to inflect a corpus of regular and irregular verbs and
lesioned it by removing connections or hidden units at
different stages in training. Marchman found that the net-
work was better able to relearn the irregular verbs than the
regular verbs, demonstrating a dissociation between these
morphological categories akin to that observed in speci-
cally language impaired (SLI) subjects. However, March-
man failed to find the opposite pattern of results, i.e., where
regulars are spared relative to irregular forms, as has been
reported in William’s syndrome (Clahsen & Almazan,
1998) and some aphasics (Marslen-Wilson & Tyler, 1997).
In fact, to our knowledge, there are no single-system mod-
els of the double dissociations between regular and irregu-
lar inflectional processing that have been found with human patients.

Elman and Hare (1997) trained a uni-modal, single-
route network to inflect phonological representations of
stems to their past tense forms. They compared activation
across the hidden units in the trained network for verbs
that were correctly inflected. Elman and Hare (1997) found
that irregular verbs produced different patterns of activa-
tion as to that observed for regular verbs, with more hidden
units active when processing irregular verbs. Elman and
Hare’s results suggest that functional specialisation does not
depend upon the presence of distinct representational
spaces in the model. On this view, functional specialisation
occurs at the level of the unit or connection, where individ-
ual units or connections may play a more prominent role in
processing, say, irregular verbs than regular verbs. This
finding suggests that even within this uni-modal, single-
route model, it might be possible to perform lesioning
experiments which selectively target regular or irregular
verbs.

3. Artifacts of scale?

Bullinaria and Chater (1995) have questioned the validity
of these connectionist accounts of double dissociations.
They argue that the observed fractionations in the behav-
iour of lesioned networks are an artifact of network scale.
In particular, they argue that large networks that have an
overcapacity of resources (e.g., number of hidden units)
relative to the problem at hand do not exhibit double disso-
ciations on lesioning. They suggest that double dissoci-
ations can only be modelled in networks when the
contribution of individual units to overall network perfor-
ance is significant. Since this is rarely the case in the
brain (they assume, reasonably, that allocation of neural
resources in the brain is massively redundant), Bullinaria
and Chater (1995) conclude that current connectionist
accounts of double dissociations do not pose a challenge
to traditional neuropsychological accounts of double
dissociations because they all involve small scale systems.

Gonnerman, Harm, and Andersen (1997) have chal-
enged Bullinaria and Chater’s position, arguing that the
failure to find double dissociations in large scale networks
is itself an artifact of not incorporating token frequency
information in the training regime. Gonnerman et al. (1997)
report that large scale networks similar to those examined
by Bullinaria and Chater but trained with the appropriate
frequency manipulations occasionally exhibit double disso-
ciations in behaviour when lesioned. The question then
remains open as to whether single-system, large-scale net-
work models of inflectional morphology can also exhibit
the type of double dissociations observed in aphasic
patients.

Plaut (1995) has also argued that individual lesions to
large scale connectionist networks can give rise to idiosyn-
cratic effects, resulting in double dissociations. He questions
the validity of such findings for our understanding of
human cognitive neuropsychology: Variability in perfor-
mance in small networks may not scale up to equivalent
lesions in the human brain. Of course, the connections or
units in a network need not necessarily correspond directly
to individual synapses or neurons in the human brain. Indeed, a common assumption of most ‘cognitive’ level
connectionist models is that individual units or connections
represent whole clusters of neurons or synapses in the real
brain. Hence, the outcome of individual lesioning experi-
ments in a connectionist network can correspond to the
outcome of a lesion to an individual patient. In this case, as
Plaut (1995) points out, it is important to compare the dis-
tribution of outcomes resulting from lesions in a broader
population of networks and patients.

In this paper, we present a large-scale, single-route
model that exhibits double-dissociations between regular
and irregular inflectional morphology. These double disso-
ciations are the result of simple variance in performance
from a stochastic norm when the network is damaged. In
this model, the effects of damage are unpredictable. Fur-
thermore, these effects may differentially affect different
words or word categories. Our experiments show, for exam-
ple, that a single network can be damaged randomly in such
a way as to have very good performance on a particular
class of words, or very bad performance, despite the level of
damage being the same in either case. We argue that the
mere observance of a double dissociation, particularly as a
rare or pathological case, is not sufficient evidence to con-
clude a separation of processing, especially in cases where
the damage itself can only be observed crudely and the
function lost is highly complex.

We present two sets of simulations to test the perfor-
ance and variations in performance of a standard single-
mechanism connectionist network that inflects noun and
verb stems of a large sample of English monosyllabic
words. In the first experiment, we test network performance
under a variety of lesion severities and determine the
expected response of the network to damage. In particular,
we examine whether noun morphology can be expected to be more or less resistant to damage than verb morphology and whether regular forms are more resistant to damage than irregular forms in these networks. The second experiment is an attempt to capture and measure expected variance of a single network’s response to very similar lesions, as a possible explanation for variable post-lesion performance as a source for varying dissociations.

4. Examining effects of lesion size

The network model we chose to damage is the same as that used by Plunkett and Juola (1999) in modelling the acquisition of past tense and plural inflection in English children. This model succeeded in simulating many aspects of inflectional development in children and achieved a high final level of performance and generalisation.

4.1. Network and training environment

The network is a standard connectionist simulation, constructed as a multi-layer perceptron network using back-propagation of error (Rumelhart, Hinton, & Williams, 1993). The simulation was built using the Stuttgart Neural Network Simulator (Zell, 1994) using 130 units for the input layer, 160 units for the output layer, and 200 units as the hidden layer. The system was trained with a learning rate of 0.1 and no momentum. Weights were initially randomised in the range ±0.5. Training was performed by a pattern update schedule, where each pattern was presented individually to the network (in random order) and training performed on each pattern. The architecture and training parameters are the same as that used by Plunkett and Juola (1999) in modelling the acquisition of past tense and plural inflection in English children.

Five random sets of starting weights were trained for 250 epochs on a corpus of stimuli encompassing 2280 noun types and 946 verb types of varying frequencies, representing their token frequencies as found in the Brown corpus (Kucera & Francis, 1967). The random seeds used to generate the five networks only affected the initial random weights. All other training characteristics, such as random order of selection of training patterns, were constant across the five networks. The training data for the simulations were taken from the CELEX corpus (Baayan, Piepenbrock, & Gulikers, 1995); we extracted from this database all words which were monosyllabic, which contained no “foreign” sounds in their pronunciation according to the Moby Pronunciator database (Ward, 1997), and for which we had evidence that they could be used as nouns or verbs. By restricting the training corpus to monosyllabic words, we are underestimating the proportion of regular inflections typewise. However, this will have little effect on the balance between regulars and irregulars tokenwise. This yielded a total corpus of 2626 stems, which encompassed 3226 total inflected types (2280 nouns and 946 verbs). Of these types, 26 were irregular nouns and 122 were irregular verbs. For these words, we took the corresponding token frequencies (of the stems) from the Brown corpus (Kucera & Francis, 1967) as a rough measure of token frequencies in running speech. The token frequencies of words were individually tabulated as nouns and verbs, and then the function \(\log_2(freq^2 + 1)\) applied to these frequencies to flatten the distribution between the high and low frequency forms. This manipulation ensured the network would have sufficient opportunity to learn low frequency irregular forms. The final variance was between 1 and 21 tokens/inflected type, meaning that the most frequent words appeared just over twenty times as often as the least. These token frequencies were also heavily dominated by nouns. Of the 17129 tokens in the training set 13045 were noun tokens (204 of them irregular) and 4084 were verb tokens (997 of them irregular). Again, this training set and frequency compression scheme was the same as that used by Plunkett and Juola (1999) in modelling the acquisition of past tense and plural inflection in English children.

The training corpus was prepared by converting the Moby symbolic pronunciation (Ward, 1997) into a large binary vector using a modification of the PGPfone alphabet representation (Juola & Zimmermann, 1996), summarised in Table 1. Each phoneme was represented as a cluster of 16 binary phonetic features including aspects such as place, manner, and height of articulation, as described in Fig. 1. Each word was divided into onset–nucleus–coda constituents and right-justified within a CCCVVCCC template (e.g., the word “cat” (/kAt/) would be represented by the training pattern ###k##A##t, where ‘#’ represents an absent sound). To this 128-bit pattern, two additional bits were appended representing the syntactic form to be inflected into, either the past tense (of a verb) or the plural (of a noun). The desired outputs were a similar encoding of the phonology of the inflected form, including an optional epenthetic vowel and final consonant.

The networks were trained using the entire log-frequency weighted corpus of nouns and verbs. The five trained networks were lesioned by randomly removing individual weights (or equivalently by setting them to zero) while leaving the basic network topology intact. The lesioning was performed at nineteen different levels, consisting of single percentages from 1% (leaving 99% of the network weights unchanged) to 10%, at 15% and at increments of 10% from 20% to 90% (leaving only 1 weight in 10). Each network was lesioned 10 times at each level, yielding a total of 950 separate experimental “participants”. Note that the connections emanating from the two input units encoding syntactic category were never lesioned. Lesioning these connections would constitute

---

1 Plunkett and Juola (1999) evaluated both incremental and mass training regimes and found that the endpoint performances of the networks were virtually identical.
depriving the network of information of the task it is supposed to perform. Our assumption is that deficits in performance arise from difficulties in processing the inflectional mappings rather than understanding the nature of the task.

4.2. Analysis

Each (lesioned) participant was presented with all distinct tokens from the training corpus and their corresponding outputs were recorded. Each output pattern was interpreted as correct if all output activations were \( \geq 0.6 \) when the target was 1 and \( \leq 0.4 \) if the target activation was 0. Every output unit had to fulfil these criteria for an output pattern to be judged correct. All other responses were coded as “incorrect”. Network performance was measured by counting the number of correct types per inflectional category and normalising (as a percentage) against the baseline performance in the undamaged, trained network.

4.3. Results

4.3.1. Baseline analysis

Baseline performance in the undamaged networks is summarised in Table 2. Performance on regular nouns and verbs was near ceiling (not less than 99%) for all networks. Likewise, performance on irregular verbs was very high (not less than 98%) for all five networks. Performance on the irregular nouns was more variable: Network 5 in expected 96% of the nouns correctly (25 out of 26 nouns), whilst Network 2 in expected 77% of the nouns correctly (22 out of the 26 irregular nouns). Table 2 also shows network performance on a set of 100 novel stems. These were the same as those used by Plunkett and Marchman (1993), 50 of which had the phonological properties characteristic (though not definitional) of irregular verbs. Generalisation of novel stems to regular plural forms (preservation of the stem and adding the appropriate plural allomorph) was consistently high (\( \geq 80\% \)) across all networks. Generalisation of novel stems to regular past tense forms varied between 50% and 74%. Note, however, that many of the novel testing data (including the indeterminate novel verbs) consisted of stems that are phonologically similar to irregular verbs. The generalisation data for novel verbs show therefore that most of the indeterminate novel stems were regularised but that a substantial proportion (\( \geq 50\% \)) of the novel stems similar to irregular verbs were regularised too. A statistical comparison of network performance across the five networks using a linear model revealed no main effect of network seed.

![Fig. 1. Composition of phonological representation.](image-url)
4.3.2. Damage analysis

Fig. 2 depicts the average network performance after damage for each combination of regularity, syntactic category, and damage level. Each data point is the mean performance for 50 participants over the subset of training data corresponding to each category (e.g., regular nouns), including 10 lesioning experiments for each of the five different networks. Performance is measured as a percentage of undamaged baseline performance to take into account the fact that regulars are closer to perfect performance than irregulars in the undamaged networks. Normalising performance against baseline levels eliminates the possibility of finding spurious effects of damage due to initial differences between the four inflectional types in the undamaged networks.

Fig. 2 shows that performance dropped off quickly when more than 1% of the connections are removed. Removal of more than 20% of the connections reduces performance to floor. However, it is also apparent that regular nouns and verbs were more resilient in the face of damage than their irregular counterparts, with irregular verbs the most sensitive to damage. Performance on novel nouns and verbs after damage closely tracked their regular counterparts, though novel performance was always inferior to the trained regular performance. Novel nouns exhibited a consistently higher level of regularisation than novel verbs even after regularisation levels have been normalised relative to baseline.

These findings were confirmed in statistical analyses of network performance in a 4-way analysis of variance, treating the base network as a between groups factor and level of damage, syntactic category, and regularity as within group factors. We have chosen to treat damage as a within group factor, since the same underlying network is exposed to multiple levels of damage. It might also be argued that each damaged network should be treated as a separate ‘subject’. A linear model analysis without any repeated measures adjustments, i.e., treating damage as a between groups factor, yielded entirely similar results to the ones reported. The single most significant predictor of network performance was the amount of damage; greater damage resulted in lower performance \( F(3,1)=2158 \). There were also significant main effects for regularity (regular forms were more resistant to damage; \( M_{\text{regular}}=0.26 \) (0.0072), \( M_{\text{irregular}}=0.13 \) (0.0051), \( F(3,1)=379 \) and for syntax (nouns were more resistant to damage; \( M_{\text{noun}}=0.23 \) (0.007), \( M_{\text{verb}}=0.17 \) (0.0058), \( F(3,1)=85 \). No significant main effect of underlying network was noted. All of the above effects were significant at \( p \ll 0.001 \). Significant interactions were noted between damage level and regularity, and between damage level and syntactic category \( F(3,1)=223, 50 \), both significant at \( p \ll 0.001 \). Note that these interactions remain intact even when floor and ceiling performances are removed from the analysis. No significant three-way (or greater) interactions were observed.

Network performance on novel nouns and verbs was compared to performance on regular nouns and verbs in a 3-way analysis of variance with level of damage, syntactic category and novelty as within group factors. All factors showed significant main effects: Greater damage resulted in lower performance \( F(2,1)=3660, p \ll 0.001 \), nouns were more robust than verbs \( F(2,1)=107, p \ll 0.001 \), regulars were more robust than novels \( F(2,1)=8.2, p < 0.01 \). There was one significant 2-way interaction between level of damage and syntactic category \( F(2,1)=30, p \ll 0.001 \). This interaction is driven by the floor effects in performance realised when more than 20% of the connections have been removed.

4.3.3. Frequency analysis

It has been argued elsewhere that type and/or token frequency can be one of the most important determining factors in connectionist learning (Plunkett & Marchman, 1991; Seidenberg & McClelland, 1989). Consequently, we would also expect high frequency forms to be more resistant to damage, irrespective of their categorisation. To
We found significant two-way interactions between syntactic category and frequency \((F(4, 1) = 34, p < 0.001)\), between syntactic category and regularity \((F(4, 1) = 71, p < 0.001)\), between regularity and frequency \((F(4, 3) = 83, p < 0.001)\), between base network and syntactic category \((F(4, 1) = 8.3, p < 0.001)\) and between base network and regularity \((F(4, 4) = 5.9, p < 0.001)\). We found significant three-way interactions between frequency, damage level and regularity \((F(4, 3) = 2.7, p < 0.05)\), and frequency, syntactic category and regularity \((F(4, 3) = 72, p < 0.001)\). Note that these interactions remain intact irrespective of whether floor and ceiling performances are included in the analysis. Indeed, the \(F\)-values reported here are those found when at least 2\% but not more than 7\% of the connections have been removed. However, a brief perusal of Table 3, as well as a detailed examination of simple effects in this ANOVA, reveal that it is the performance on the low frequency irregular nouns after lesioning that is driving the main effect of frequency and the three-way interactions.

### 4.4. Discussion

The lesioning simulations reported in this experiment constitute an investigation of acquired language disorder. The removal of connections from the network is intended to parallel cortical lesions observed in an hypothetical aphasic patient. It has often been reported that damage to parallel distributed processing systems results in a graceful degradation in performance (for example, see Rumelhart et al., 1986). The networks analysed in these simulations also exhibit this characteristic continuous deterioration in performance. The networks continue to perform rather well even when several thousand connections have been removed from the system. The statistical main effect of damage indicates that lesioning these networks causes a general deterioration in performance, and that the greater the degree of damage, the greater the degree of deterioration in performance. It is apparent from Fig. 2, however, that degradation in performance does not vary in a linear fashion with amount of damage. There is a limited window of damage (2–10\%) where performance drops off dramatically, indicating that a critical mass of connections needs to remain intact in order to sustain near-normal behaviour. This finding mirrors the acquisition literature (Marchman & Bates, 1994; Plunkett & Juola, 1999; Plunkett & Marchman, 1993), where it is reported that a critical mass of vocabulary items is required to trigger normal levels of generalisation in children and connectionist networks. By implication, the sudden deterioration in behaviour that is sometimes observed in acquired disorders, such as Alzheimer’s disease, may be the result of a gradual and continuous atrophy of the neural systems underpinning a specific cognitive or linguistic function. Sudden loss of behaviour need not correspond to a sudden loss of neural resources.

Another important finding is that random damage to these single-route, feed-forward networks results, on average, in differential impairments to sub-categories of the task on which the networks have been trained. In particular, irregular inflections are more likely to suffer deterioration in performance than regular inflections. Likewise, verb inflections are less robust in the face of damage than noun inflections. These results demonstrate that simple dissociations in behaviour (e.g., loss of irregular inflection but not regular inflection, or loss of verb inflection but not noun inflection) are entirely compatible with single-system approaches to regular and irregular noun and verb morphology. The observation of simple dissociations in inflectional morphology, though compatible with modular accounts (for example, Pinker & Prince, 1988), does not constitute evidence for such accounts (see also Bishop, 1997, p. 912).

An important factor determining network performance on individual forms is token frequency. This is reflected in the effects of frequency reported in the analyses above: In particular, high frequency irregular nouns are more robust to damage than low frequency irregular nouns. This result indicates that token frequency plays an important role in sculpting the organisation of the network. Training items with high token frequency encourage the formation of strong connections. In this training environment, irregular nouns constitute the inflectional class with the greatest variation of token frequency. The impact of frequency on

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Network performance on regular and irregular nouns and verbs as a function of frequency level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency quartile</td>
<td>Nouns</td>
</tr>
<tr>
<td></td>
<td>1st</td>
</tr>
<tr>
<td>Reg</td>
<td>30.9</td>
</tr>
<tr>
<td>Irreg</td>
<td>18.8</td>
</tr>
<tr>
<td>Verbs</td>
<td></td>
</tr>
<tr>
<td>Reg</td>
<td>25.7</td>
</tr>
<tr>
<td>Irreg</td>
<td>10.6</td>
</tr>
</tbody>
</table>

Percentage means reflect averages over networks and levels of damage.

examine this possibility, we divided the training set into frequency quartiles (approximately 25\% of types in each quartile, irrespective of syntactic category or regularity) and analysed network performance on the four inflectional categories for each frequency quartile. Mean performances (collapsing over networks and levels of damage) across category type and frequency level are tabulated in Table 3 which shows that irregular nouns exhibited frequency effects after lesioning. The other three inflectional categories showed stable performance across frequency quartiles. It should be noted that irregular nouns show the greatest frequency variation in the training environment.

These findings were confirmed in a 5-way ANOVA with base network as a between groups factor, and level of damage, syntactic category, regularity and frequency as within group factors. In this reanalysis, all the factors produced significant main effects: Damage level \((F(4, 1) = 531)\), frequency quartile \((F(4, 3) = 103)\), base network \((F(4, 4) = 60)\), syntactic category \((F(4, 1) = 509)\) and regularity \((F(4, 1) = 4927)\) produced significant main effects \((p < 0.001\) in all cases). We also found significant two-way interactions between syntactic category and regularity \((F(4, 1) = 34, p < 0.001)\), between syntactic category and frequency \((F(4, 3) = 71, p < 0.001)\), between regularity and frequency \((F(4, 3) = 83, p < 0.001)\), between base network and syntactic category \((F(4, 4) = 8.3, p < 0.001)\) and between base network and regularity \((F(4, 4) = 5.9, p < 0.001)\). We found significant three-way interactions between frequency, damage level and regularity \((F(4, 3) = 2.7, p < 0.05)\), and frequency, syntactic category and regularity \((F(4, 3) = 72, p < 0.001)\). Note that these interactions remain intact irrespective of whether floor and ceiling performances are included in the analysis. Indeed, the \(F\)-values reported here are those found when at least 2\% but not more than 7\% of the connections have been removed. However, a brief perusal of Table 3, as well as a detailed examination of simple effects in this ANOVA, reveal that it is the performance on the low frequency irregular nouns after lesioning that is driving the main effect of frequency and the three-way interactions.

4.4. Discussion

The lesioning simulations reported in this experiment constitute an investigation of acquired language disorder. The removal of connections from the network is intended to parallel cortical lesions observed in an hypothetical aphasic patient. It has often been reported that damage to parallel distributed processing systems results in a graceful degradation in performance (for example, see Rumelhart et al., 1986). The networks analysed in these simulations also exhibit this characteristic continuous deterioration in performance. The networks continue to perform rather well even when several thousand connections have been removed from the system. The statistical main effect of damage indicates that lesioning these networks causes a general deterioration in performance, and that the greater the degree of damage, the greater the degree of deterioration in performance. It is apparent from Fig. 2, however, that degradation in performance does not vary in a linear fashion with amount of damage. There is a limited window of damage (2–10\%) where performance drops off dramatically, indicating that a critical mass of connections needs to remain intact in order to sustain near-normal behaviour. This finding mirrors the acquisition literature (Marchman & Bates, 1994; Plunkett & Juola, 1999; Plunkett & Marchman, 1993), where it is reported that a critical mass of vocabulary items is required to trigger normal levels of generalisation in children and connectionist networks. By implication, the sudden deterioration in behaviour that is sometimes observed in acquired disorders, such as Alzheimer’s disease, may be the result of a gradual and continuous atrophy of the neural systems underpinning a specific cognitive or linguistic function. Sudden loss of behaviour need not correspond to a sudden loss of neural resources.

Another important finding is that random damage to these single-route, feed-forward networks results, on average, in differential impairments to sub-categories of the task on which the networks have been trained. In particular, irregular inflections are more likely to suffer deterioration in performance than regular inflections. Likewise, verb inflections are less robust in the face of damage than noun inflections. These results demonstrate that simple dissociations in behaviour (e.g., loss of irregular inflection but not regular inflection, or loss of verb inflection but not noun inflection) are entirely compatible with single-system approaches to regular and irregular noun and verb morphology. The observation of simple dissociations in inflectional morphology, though compatible with modular accounts (for example, Pinker & Prince, 1988), does not constitute evidence for such accounts (see also Bishop, 1997, p. 912).

An important factor determining network performance on individual forms is token frequency. This is reflected in the effects of frequency reported in the analyses above: In particular, high frequency irregular nouns are more robust to damage than low frequency irregular nouns. This result indicates that token frequency plays an important role in sculpting the organisation of the network. Training items with high token frequency encourage the formation of strong connections. In this training environment, irregular nouns constitute the inflectional class with the greatest variation of token frequency. The impact of frequency on
performance in this network is commensurate with earlier connectionist models of psycholinguistic function (Daugherty & Seidenberg, 1992; Ellis & Schmidt, 1998; Plunkett & Marchman, 1991; Seidenberg & McClelland, 1989) and the assumption that higher frequency forms are more robust to damage in humans (Bishop, 1997; Shallice, 1988).

The main effect of regularity (regulars are more robust than irregulars) and the syntactic category by frequency interaction (irregular nouns are more sensitive to frequency than irregular verbs in the face of damage) offer novel empirical predictions for acquired disorders of inflectional systems. Some selected samples of disordered populations, such as patients suffering from Alzheimer’s disease, show greater damage to irregular processing of past tense inflections (Ullman et al., 1997). However, we know of no systematic comparisons of regular and irregular performance in a representative sample of the aphasic population. The main effects and interactions reported in the current simulations clearly need further evaluation in the context of acquired disorders of inflectional morphology.

5. Exploring variability
5.1. Methods

To evaluate the degree of variability in network performance, we carried out 13,720 separate lesioning experiments, all at the 1% level (in other words, leaving a random 99% of the connections intact) on one of the networks (Network 5 in Table 2) developed in the prior experiment. Using just one network eliminates any variability that might be introduced by different random seeds at the outset of training. In this respect, our evaluation of performance variability is somewhat conservative compared to “real world” variation. Hence, the incidence of dissociations resulting from purely stochastic variation is also likely to be an underestimate of what we might expect to observe from real world situations. We present the results of the network damage at the 1% level because pilot simulations indicated that higher levels of damage resulted in ‘floor’ effects for irregular inflections, thereby precluding the possibility of observing double dissociations.

Performance was evaluated after the network had been trained for 250 epochs and then lesioned. Note that prior to lesioning, Network 5 performs at a very high level (again see Table 2). Also note that the connections emanating from the two input units encoding syntactic category were never lesioned. Individual inflected forms were diagnosed as being correct or incorrect in the same manner as in the baseline simulations. In a medical context, this might represent an exploration of the range of performance that could be expected after a particular patient received an injury of some particular severity, but on nearly 14,000 successive occasions (assuming that the damage could be undone each time).

We repeated this set of 13,720 lesioning experiments with another network (Network 1 in Table 2) developed in the prior experiment, to determine whether the same variability of performance would be observed. Note that we analyse the variability occurring within each of these networks separately, not across the networks.

5.2. Results

Table 4 lists measures of network performance after damage on each inflectional type as a percentage of baseline performance for all 13,720 lesioned networks for Network 5. As we would expect from the previous set of lesioning experiments reported in the baseline simulations, regular inflections are more robust to damage than irregular inflections and nouns are more robust to damage than verbs. However, there is considerable variation in performance across the 13,720 lesioned networks within an inflectional class, despite the fact that the same level of lesioning (1%) has been performed in each case. Even though variation in patient (random start seed) and physical severity of damage have been controlled for, performance exhibits a positively skewed normal distribution. For some lesioned networks, performance remains at 100% (relative to baseline) for a given inflectional type even after 580 of the connections have been removed whereas for other networks performance deteriorates to less than 60% of baseline. Clearly, some of the connections in the network play a more important role in the maintenance of performance than others.

The question arises as to whether lesioned networks that are particularly bad (or good) on one inflectional class are likewise particularly bad (or good) on the other inflectional classes. In other words, does a particular type of damage to the network target different inflectional categories to the same extent? In terms of network architecture, does the removal of specific sets of connections cause deterioration in performance across inflectional classes or does the removal of specific sets of connections target performance on some inflectional classes but not others?

Fig. 3 depicts four scatterplots comparing network performance on pairs of inflectional classes. Each point on a scatterplot represents a single network’s performance (after lesioning) on the inflectional pair. Each scatterplot contains the relevant inflectional comparison for every lesioning experiment. Fig. 3a compares performance on irregular verbs and irregular nouns. The plot reveals a

---

Table 4

<table>
<thead>
<tr>
<th>Inflectional Type</th>
<th>Network 5 Performance</th>
<th>Network 1 Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>Verbs</td>
<td>Novel</td>
</tr>
<tr>
<td>Reg</td>
<td>Irreg</td>
<td>Reg</td>
</tr>
<tr>
<td>Mean</td>
<td>96.5</td>
<td>76.2</td>
</tr>
<tr>
<td>SD</td>
<td>3.4</td>
<td>14.8</td>
</tr>
</tbody>
</table>

2 Some points overlap, so there are not 13,720 distinguishable results visible on each scatterplot.
A considerable degree of scatter indicating that irregular noun performance is not a particularly good predictor of irregular verb performance. Indeed, statistical analyses indicate that there is a rather low correlation in performance between irregular verbs and irregular nouns ($R^2 = 0.18$). In general, the level of correlation in performance across inflectional classes reflects the degree to which different inflections share the same network resources. Fig. 3b compares performance on irregular and regular nouns. The data points in this scatterplot populate the upper half of the figure, reflecting the relative robustness of regular nouns compared to irregular nouns. The correlation in performance between regular and irregular nouns is also quite low ($R^2 = 0.19$), suggesting that irregular nouns share network resources with regular nouns to a similar extent as irregular verbs. Fig. 3c compares performance on regular and irregular verbs. Again, the data points populate the upper half of the scatterplot, reflecting the relative robustness of the regular verbs. The correlation in performance between regular and irregular verbs is very low ($R^2 = 0.05$) indicating that they share network resources only to a limited extent. Finally, Fig. 3d compares performance on regular nouns and verbs. The robust performance on these regular types results in a high degree of clustering in the upper righthand quadrant of the scatterplot. Performance on these inflectional types also shows a moderate level of correlation ($R^2 = 0.26$).

### 5.2.1. Identifying double dissociations

Given the variability in the distribution of performance when comparing inflectional classes, one might expect variation in the likelihood of finding dissociations in performance between them. For example, the distribution of points in Fig. 3b suggests that it should be quite easy to find lesioned networks where regular nouns exhibit intact performance while irregular noun performance is impaired (upper left-hand quadrant of Fig. 3b) but not possible to find examples of lesioned networks where irregular noun performance is intact while regular noun performance is impaired (lower right-hand quadrant of Fig. 3b). In contrast, the greater scattering of points in Fig. 3a suggests that it might be possible to find examples of dissociations between irregular nouns and irregular verbs in both directions, i.e., a double dissociation.

Identification of a double dissociation implies an accepted definition of what constitutes intact and impaired performance. The neuropsychological literature does not offer an agreed, statistical definition of double dissociation (Shallice, 1988, p. 239). Neuropsychologists do agree that a double dissociations exists when patient X shows 100% performance on
task A and 0% performance on task B, while patient Y shows 0% performance on task A and 100% performance on task B. Unfortunately, such contrasts on theoretically interesting comparisons rarely (if ever) occur and it is more usual to categorise patients as falling within or outside the “normal range” and identify double dissociations on this basis. However, it is often unclear when levels of performance on a task are considered to be outside the normal range.

Since we have a large number of “subjects” in our simulation experiments, we can explore alternative statistical definitions of what constitutes intact and impaired performance and examine the implications for identification of double dissociations. We will examine two types of definition: First, we will define impairment (and intact performance) in terms of absolute percentages of baseline performance. Second, we will define impairment (and intact performance) in terms of the degree of statistical variance from a measured mean. In each case, we tabulate the pattern of dissociations in performance observed between the inflectional classes modelled in our networks.

Let us define impaired performance as ≤ 50% of baseline performance and intact performance as ≥ 95% of baseline performance. A dissociation between two tasks occurs when a lesioned network correctly inflects ≤ 50% of one inflectional class but simultaneously inflects ≥ 95% of another inflectional class. Call this a stringent definition of dissociation. Alternatively, we can define impaired performance as ≤ 70% of baseline performance and intact performance as ≥ 90% of baseline performance. Call this a relaxed definition of dissociation. Table 5 lists the number of dissociations that occur across all 13,720 lesioning experiments when these stringent and relaxed definitions of intact and impaired performance are applied.

For both of these sets of definitions of intact and impaired performance, the overwhelming majority of dissociations involve intact performance on regular forms with impaired performance on irregular forms. For the stringent definition, there is just one network in which the regularity effect is reversed, i.e., the case where irregular verbs are intact but regular verbs are impaired. This one case together with the large number of cases (126) where regular verbs are intact and irregular verbs are impaired provides us with the first evidence of double dissociation in this single-route network. Notice that there is no double dissociation between regular and irregular nouns. Nor do double dissociations occur between regular nouns and verbs. It is, however, possible to find evidence of double dissociation between irregular nouns and verbs.

Applying the relaxed definition of intact and impaired performance, the number of dissociations increases dramatically. However, the overall pattern of results does not. The overwhelming majority of dissociations still involve intact performance on regular forms and impaired performance on irregular forms. Double dissociations are observed between irregular nouns and verbs. Most importantly, double dissociations are observed between regular and irregular verbs but not between regular and irregular nouns.

Percentage-based definitions of intact and impaired performance do not take into account the mean and the distribution in performance across networks observed for different inflectional classes. However, it is precisely the distribution in performance across a population that the cognitive neuropsychologist might find useful in determining a definition of intact and impaired behaviour. Unfortunately, distributional information is rarely taken into account:

No single-case study in which the patient was selected on the basis of a dissociation has, …, used any form of statistical justification of the selection procedure. (Shallice, 1988, p. 239)

Suppose for each performance metric (e.g., irregular noun tokens, regular verb tokens, etc.), we fit a Gaussian distribution to performance levels across all 13,720 lesioning experiments to define a mean and standard deviation. We use the mean of the lesioned networks rather than baseline performance simply to reflect the fact that double dissociations are identified in populations which suffer some kind of abnormality. In these circumstances, intact performance will be that of a system that has suffered some kind of damage. Furthermore, performance distributions have a strong positive skew reflecting the relative robustness of these distributed information processing systems in the face of damage. Note that networks with levels of performance at least two standard deviations above the mean all operate within the normal range. We consider a lesioned network to be intact if its performance is at least two standard deviations above the mean (or is better than 95%), and impaired if its performance is at least two standard deviations below the mean (see Bates, Applebaum, & Allard (1991) for a similar definition).

Using the means and standard deviations for each inflectional category reported in Table 4, Table 5 tabulates the frequency of simple dissociations between pairs of inflec-
tional types when a variance-based definition of dissociation is applied. The number of simple dissociations observed is quite similar to that observed in the stringent definition of dissociation. However, the overall pattern of results changes in some important respects. First, double dissociations are now observed between regular and irregular nouns where none were found before. Likewise, regular nouns and verbs doubly dissociate where they did not before. Surprisingly, irregular nouns and verbs no longer doubly dissociate. The pattern of double dissociation between regular and irregular verbs is preserved.

The disappearance of double dissociations between irregular nouns and verbs is due to the high degree of variance in performance in these categories (see Table 4). On this variance based definition of dissociation, irregular nouns are impaired when their performance drops below 46.6% of baseline. Although there are many lesioned networks where this occurs (see Fig. 3a), there are no networks where irregular verb performance is simultaneously at a level \( \geq 95\% \). In contrast, the appearance of double dissociations between regular nouns and verbs is triggered by the relatively low variance in performance in these classes. Impaired performance on regular nouns is now defined as \( \leq 89.7\% \) of baseline and on regular verbs as \( \leq 75.6\% \) of baseline. It is now relatively easy to find examples of intact regular verb performance with simultaneously impaired regular noun performance, whereas before such dissociations were non-existent. Hence, double dissociations between regular nouns and verbs are easy to find.

The incidence of dissociations between novel forms and their regular and irregular counterparts (for all three definitions of dissociation) is also listed in Table 5. This analysis shows that when a stringent definition of intact and impaired performance is used, no double dissociations are observed between novel forms and their regular or irregular counterparts. Only dissociations where novel forms are preserved relative to their irregular counterparts are observed. When the definition of intact and impaired performance is relaxed, the number of simple dissociation increases dramatically, resulting in multiple double dissociations between novel and irregular verbs. Finally, the distribution of double dissociations switches from verbs to nouns when a variance based definition of intact and impaired performance is used. Novel nouns doubly dissociate from both regular and irregular nouns on this definition.

5.3. Control analyses

Four factors contribute to the variability in performance after lesioning across the inflectional classes.

1. Algorithmic differences in the inflection of regular and irregular forms.
2. The token frequency of the individual words that constitute an inflectional class.
3. The class size (or type frequency) of a given inflectional category.
4. Random factors deriving from the specifics of the training and lesioning procedures (random start seed, order of training materials, and random selection of lesion sites).

We investigated the contribution of algorithmic differences to performance variability by creating a set of pseudo-categories in which word stems were assigned to an inflectional class in a random fashion rather than on the basis of their mapping type, e.g., add a suffix, or change a vowel. Using the same set of 13,720 lesioning experiments, we assigned word stems to categories randomly, preserving the original size of each inflectional class. Consequently, pseudo-irregular verbs consisted of a mixture of regular and irregular nouns and verbs, as did all the other pseudo-categories. We evaluated performance on the basis of each word’s original inflectional target. If algorithmic differences in the inflection of regular and irregular forms do not contribute to variability in performance after lesioning we should expect to see the same distribution in performance as reported in Table 4. Table 6 lists measures of network performance after damage on each pseudo-inflectional type as a percentage of baseline performance for all 13,720 lesioned networks.

The overall distribution of performance is almost identical across the four pseudo-inflectional classes, in contrast to that presented in Table 4. This result confirms the claim that algorithmic differences in regular and irregular verb inflection have an important impact on performance variability across classes. In comparison to the original analyses presented in Table 4, the primary differences are between the two irregular categories. Whereas irregular verbs and nouns had standard deviations of 14.5 and 14.8, respectively, the pseudo categories have corresponding standard deviations of 3.8 and 5.6. Note that each pseudo class has the same size as the corresponding original class. Note also that pseudo irregular nouns, the smallest class, show the highest standard deviation in performance (5.6). This result supports the claim that class size has an important impact on performance variability.

Likewise, we can consider the impact of algorithmic differences in inflectional type on the likelihood of finding dissociations in performance, by comparing dissociations between pseudo categories. Again, if algorithmic differences in the inflection of regular and irregular forms do not contribute to the likelihood of finding dissociations in performance we should expect to see the same distribution in performance as reported in Fig. 3.

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Performance for pseudo inflectional types (percentage of baseline) in Network 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo nouns</td>
<td>Pseudo verbs</td>
</tr>
<tr>
<td>Regular</td>
<td>Irregular</td>
</tr>
<tr>
<td>Mean</td>
<td>93.8</td>
</tr>
<tr>
<td>SD</td>
<td>3.7</td>
</tr>
</tbody>
</table>
Fig. 4 shows the performance of 13,720 networks on four pairs of pseudo categories, as well as a statistical measure ($R^2$) of the correlation in performance between pseudo classes. The distribution of points in the four scatterplots in Fig. 4 is quite different to the corresponding scatterplots in Fig. 3. In general, the plots exhibit a much tighter clustering in the control analyses. This is reflected in much higher levels of correlation in performance: Performance on pseudo regular verbs correlates almost perfectly with performance on pseudo regular nouns ($R^2 = 0.95$) and even pseudo regular verbs and pseudo irregular verbs (which were very poorly correlated in the original comparisons) are now highly correlated ($R^2 = 0.73$). Comparisons involving pseudo irregular nouns show a greater degree of scatter and the correlations involving pseudo irregular nouns are correspondingly lower. It is noteworthy that pseudo irregular nouns constitute the smallest class in these analyses. Nevertheless, their correlations with pseudo regular nouns and pseudo irregular verbs are quite high compared to the original analyses. These control analyses demonstrate that when the algorithmic differences between inflectional categories has obliterated the occurrence of double dissociations. The only circumstance where simple and double dissociations occur is for the variance based definition of dissociation, and here, only for comparisons involving pseudo irregular nouns, reflecting the impact of class size on the likelihood of observing a dissociation.

5.4. Network 1

Table 7 lists measures of network performance after damage on each inflectional type as a percentage of baseline performance for all 13,720 lesioned networks for Network 1. The pattern of performance after damage is similar to Network 5 (see Table 4), with a clear advantage
for regulars over irregulars. The noticeable differences are relatively higher levels of performance on irregular verbs and lower performance on novel nouns.

Table 8 lists measures of network performance after damage on each pseudo-inflectional type as a percentage of baseline performance for all 13,720 lesioned networks. Again, the pattern of performance is very similar to Network 5 (see Table 6).

Finally, Table 9 lists the number of dissociations that occur across all 13,720 lesioning experiments when the stringent, relaxed and variance based definitions of intact and impaired performance are applied to Network 1. The pattern of dissociations occurring between the inflectional classes is similar to that observed for Network 5: Depending on the definition used, double dissociations are observed between all the extant inflectional classes, and double dissociations between novel classes and trained classes occur across all four comparisons, not just three. The only cases, where double dissociations do not occur in Network 1 is in the Control comparisons. Here, it will be recalled that Network 5 yielded a double dissociation between irregular nouns and verbs. No such double dissociation was observed in Network 1. This variability across networks is to be expected given the small class sizes of the control irregulars.

The findings for Network 1 highlight the replicability of the lesioning experiments with Network 5.

Table 8

<table>
<thead>
<tr>
<th></th>
<th>Pseudo nouns</th>
<th>Pseudo verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regular</td>
<td>Irregular</td>
</tr>
<tr>
<td>Mean</td>
<td>92.95</td>
<td>92.76</td>
</tr>
<tr>
<td>SD</td>
<td>3.85</td>
<td>5.86</td>
</tr>
</tbody>
</table>

Table 9

The number of networks showing simple dissociations between inflectional categories for alternative definitions of intact (+) and impaired (−) performance in Network 1.

5.5. Discussion

This second set of simulations was performed to examine whether double dissociations between inflectional types could be found when a uni-modal, single-route model of noun plurals and verb past tense morphology was lesioned in a random fashion. Given the statistical definition of double dissociation adopted here, we have observed that double dissociations exist in all comparisons of inflectional types. Of particular importance is the finding that regular and irregular past tense inflections doubly dissociate from one another. Likewise, regular and irregular noun plural inflections also doubly dissociate from one another. The observation of such double dissociations in aphasic patients has traditionally been interpreted as evidence for a modular account of the mental representation of inflectional morphology (Pinker, 1991). The current work demonstrates that these types of double dissociations can also be found in uni-modal, single-route models. Furthermore, the general pattern of findings are replicable in different networks.

We have also demonstrated that double dissociations occur between noun and verb morphology in a single-route model. It will be recalled from the first set of simulations that noun inflections were more robust in the face of damage than verb inflections. Our findings predict that although the simple dissociation of verb inflection from noun inflection should be more commonly observed in aphasic populations, the opposite pattern of simple dissociation will also occur, albeit more rarely. Again, we suggest that the identification of such double dissociations, though compatible with a modular account (nouns versus verbs), is not evidence for such an account, since precisely the same effects can be observed in a single-route model. Indeed, the relative likelihood of observing different types of simple dissociation might offer a way to distinguish the different theoretical accounts: Modular accounts are not committed to one type of dissociation being more common than any other, whereas single-route accounts are.

There are several features of the profile of double dissociations observed in these single-route models that are worth highlighting. First, double dissociations in the unimodal model presented here are exceedingly rare. This follows directly from the fact that different inflectional types share the same processing resources, so that lesioning connections affects performance across the board (see Fig. 2). Hence, we observe a generally high correlation in performance loss after lesion between inflectional types. The implication of this finding is clear: If double dissociations between theoretically significant categories are commonly observed across aphasic patients, then it is likely that they derive from the malfunction of functionally separable mechanisms. However, if such double dissociations are rarely observed then it is entirely possible that they derive from stochastic variation in performance resulting from damage to a single mechanism. Under these circumstances, the identification of a double dissociation in performance
on two tasks across participants is not a good basis for inferring a separation in the mechanisms underlying those tasks.

Second, we showed in the control simulations that small class size was a significant contributing factor to variability in network performance, and therefore, to the likelihood of observing double dissociations. Likewise, the use of test batteries with only a few items per class when evaluating aphasic patients is likely to lead to greater variability in performance across patients and hence exaggerate the likelihood of observing double dissociations.

Third, the probability of observing double dissociations between inflectional types varies. This is most clearly demonstrated by inspecting the scatterplots in Fig. 3a–d. For example, double dissociations between irregular verbs and irregular nouns are comparatively easy to find whereas double dissociations between irregular verbs and regular verbs are few. Furthermore, although double dissociations between regular and irregular verbs are relatively few, they are more frequent than double dissociations between regular and irregular nouns. This variability constitutes an empirical prediction about the likelihood of observing specific dissociations in performance across the aphasic population. Indeed, an extreme prediction of these simulations is that there should be no double dissociations between regular and irregular nouns when double dissociations are defined in absolute terms. To our knowledge, no one has ever reported such double dissociations.

This variability in performance between inflectional types also offers some insights into the degree to which processing resources are shared by noun and verb inflections in these single-route models. Lesions that target connections which share equal responsibility for processing different inflectional types will, by definition, result in similar decrements in performance for those types. The greater the division of labour between connections for processing different inflections, the greater the variation in the effects of lesion between those types. For example, the lesioning of an hypothetical connection which has no responsibility for processing regular verbs would have no effect on regular verb performance. Hence, the level of correlation between performance on different inflectional types after lesion offers a first approximation of the degree to which processing resources are shared in these models.

Not unexpectedly, random classifications of the training set (as assessed in the control analyses) showed the highest levels of correlation in performance after lesion. Given the arbitrary character of this classification, there is no reason to suppose that connections in the network will have become specialised with respect to these groupings. In contrast, the correlation in performance between regular and irregular verbs is quite low, indicating a higher degree of specialisation in processing resources. The correlation between regular and irregular nouns indicates an intermediate level of specialisation. We observe, then, that the degree to which individual connections become functionally specialised as a result of training varies along a continuum.

The low level of correlation in performance between regular and irregular verbs highlights the capacity of single-route models to exhibit an emergent functional specialisation. The results demonstrate that the degree of functional specialisation in the network varies with the type of mapping that the network has to learn and the distributional characteristics of the training set. For example, regular and irregular nouns show much less functional specialisation than regular and irregular verbs. Indeed, it was more difficult to find double dissociations between noun inflections than verb inflections (see Table 5). In contrast, a modular approach to noun and verb inflection would predict that both grammatical categories should exhibit clearly dissociable patterns of impairment between regular and irregular forms, depending on the locus of the brain damage (Pinker, 1991).

It is also worth noting here that the scale of the network used in the current simulations is much larger than the networks in earlier models of double dissociations. For example, the Plaut & Shallice (1994) model has approximately 5000 connections while the current model has 58,000 connections. Likewise, the size of the training set in the current model is several orders of magnitude larger than the vocabulary set used in the Plaut & Shallice (1994) model (40 words versus 3226 words). It is unlikely, therefore, that the pattern of results reported here is an artifact of small network scale or size of testing samples.

It is important to emphasise the conservative approach we have taken in our attempt to model double dissociations in a single-route model. Most importantly, the models used only unimodal phonological representations and double dissociations were only sought through damage to single source networks. Recently, McClelland & Patterson (2002) have argued that the phonological complexity of regular verbs make them more susceptible to impairment than irregular verbs. Hence, they point out that when phonological complexity is controlled for in the testing materials used with aphasic patients, damage to phonological processing pathways has similar effects on regular and irregular verbs. The current set of simulations used monosyllabic forms for all inflectional classes. Furthermore, the feed-forward network used to model the inflectional mapping is equally adept at producing long outputs as it is at producing short outputs: The suffixation process associated with regular inflections does not present any more additional complexity than does an irregular vowel change, in the context of the current architecture (see Plunkett & Marchman (1991) for a detailed discussion of this issue). Hence, phonological complexity is not a source of variability in the current simulations. However, the introduction of multiple levels of phonological complexity to the training materials (and the use of recurrent architectures) in these simulations would serve to increase the variability of performance within classes, and therefore, the likelihood of observing double dissociations across inflectional classes.
6. Conclusions

Our lesioning experiments have shown that damage to distributed information processing systems can result in selective impairment in performance of some, but not other tasks represented in the system. These findings offer further support to previous work showing that dissociations in behaviour can result from damage to single-route, distributed processing systems (Bullinaria & Chater, 1995; Farah & McClelland, 1991; Gonnerman et al., 1997; Marchman, 1993; Plaut, McClelland, Seidenberg, & Patterson, 1996; Plaut & Shallice, 1994).

The first set of lesioning experiments (baseline simulations) demonstrated that the effects of damage depended on:

- The number of connections removed: The greater the number of connections removed, the greater the deficit in performance. We observed a non-linear, sigmoidal relation between damage and level of performance.
- The task evaluated: Nouns were generally more robust than verbs and regulars were more robust than irregulars.
- The frequency of the trained items: High frequency irregular nouns were more robust to damage than low frequency irregular nouns.

The second set of lesioning experiments demonstrated that the effect of lesioning on performance varies stochastically: The population of lesioned networks exhibit a positively skewed normal distribution in levels of performance with most networks performing quite well but below normal, a few performing within the normal range, and a few performing extremely badly. Deficits in performance after lesion are highly correlated across different tasks: Networks that tend to be better (or worse) than average on some task, say regular verbs, tend to be better (or worse) than average on other tasks, say irregular verbs. However, this correlation is not perfect. Consequently, it is possible to find some networks that show simple dissociations between tasks, say good performance on irregular verbs but poor performance on regular verbs, even after quantitatively identical amounts of lesioning. Importantly, these simple dissociations are observed between all types of inflectional processes, giving rise to double dissociations.

According to the “logic of cognitive neuropsychology” (Coltheart, 1985), double dissociations can be used to diagnose separation in the mechanisms underlying the two tasks. Our results indicate that the occurrence of a double dissociation does not logically require the inference that separate underlying mechanisms are involved. Uni-modal, single-route accounts of inflectional morphology predict that double dissociations will occur between regular and irregular nouns and verbs, merely as a result of stochastic variation in performance after lesion in a distributed processing system. The uni-modal, single-route account also predicts that the incidence of double dissociations should be very low in any aphasic population. Typically, inflectional impairments should be system wide as opposed to being restricted to a given inflectional type. Given the rarity of double dissociations in the aphasic population (Parkin, 1996), against a background of large numbers of individuals presenting with developmental and acquired disorders, we may conclude that stochastic interpretation of double dissociations that do not infer separable mechanisms, may have wider applicability than normally supposed.

It is important to note that we do not claim that double dissociations have no role to play in identifying the manner by which cognitive processes are fractionated, either functionally or neuroanatomically. The stochastic interpretation of double dissociations offered here is not meant to imply that single-route models apply to all aspects of cognition! Indeed, we have highlighted the prediction that a more even distribution of double dissociations between inflectional processes in the aphasic population would count against a uni-modal, single-route account. Our primary claim is that the observation of isolated examples of double dissociations (even of the classic variety) do not mean much by themselves.

Again, it is important to clarify that the definition of double dissociation used here is purely behavioural (see the definition quoted from (Shallice, 1988) at the beginning of this paper). Traditionally, the identification of double dissociations has included the localisation of separate neuroanatomical lesions putatively responsible for the pattern of deficits observed (see Teuber, 1955; Weiskrantz, 1968). When accompanied by evidence of non-overlapping lesions, the inference of separable mechanisms from behavioural double dissociations carries added weight.

We have presented a model and explanation for some kinds of double dissociation that does not require a different method of processing or even a functional separation between modules in the underlying processing. We argue instead that, because the effects of damage to a system as complex as inflectional morphology are somewhat unpredictable, in some cases “random” damage will result in surprisingly good performance on some aspects of a task and surprisingly bad performance on other aspects. Small differences in representation may occasionally conspire (under damage) to produce variance among some inflectional groups, leading to the observation of a double dissociation.

These findings caution against the extensive overreliance on individual case studies in the cognitive neuropsychological literature. By sorting through enough aphasic patients, one can expect to discover outliers that may be woefully misleading in the absence of some analysis of how characteristic they are for the aphasic population as whole.

Acknowledgments

This research was supported by an ESRC project grant to the first author, and a WelIcome Trust studentship to the second author. Thanks also to Patrick Juola who carried out some pilot simulations reported in Juola & Plunkett (2000).
References


