Interruptibility as a Constraint on Hybrid Systems[*]

Richard Cooper

Department of Psychology
University College
Gower Street
London WC1E 6BT

email: ucjtrrc@ucl.ac.uk
telephone: +44 71 387 7050 (Ex. 5372)
fax: +44 71 436 4276

Bradley Franks

Department of Social Psychology
London School of Economics
Houghton Street
London WC2A 2AE

email: franks@lse.vax.ac.uk
telephone: +44 71 955 7704
fax: +44 71 955 7565

Keywords:

Cognitive models; Interruptions; Hybrid Systems; Connectionism;
Symbolic Systems; Constraints on Hybrid Systems

Abstract

It is widely mooted that a plausible computational cognitive model should involve both symbolic and connectionist components. However, sound principles for combining these components within a hybrid system are currently lacking: the design of such systems is often *ad hoc*. In an attempt to ameliorate this we provide a framework of types of hybrid systems and constraints therein, within which to explore the issues. In particular, we suggest the use of “system independent" constraints, whose source lies in general considerations about cognitive systems, rather than in particular technological or task based considerations. We illustrate this through a detailed examination of an interruptibility constraint: handling interruptions is a fundamental facet of cognition in a dynamic world. Aspects of interruptions are delineated, as are their precise expression in symbolic and connectionist systems. We illustrate the interaction of the various constraints from interruptibility in the different types of hybrid systems. The picture that emerges of the relationship between the connectionist and the symbolic within a hybrid system provides for sufficient flexibility and complexity to suggest interesting general implications for cognition, thus vindicating the utility of the framework.

1 Constraints on Hybrid Systems

In attempting to design general, cognitively plausible computational systems, many would contend that the cognitive sciences have reached a consensus, which is also an impasse. Whilst there is an acknowledged necessity for hybrid systems incorporating techniques of both symbolic and connectionist systems,

[*] We would like to thank Jean Aitchison, Nick Braisby, John Fox, David Green, Celia Hayes, Hans van der Koot, John Morton, Tim Shallice, Andy Wells and two anonymous referees for helpful discussions and comments that have contributed, directly and indirectly, to the ideas presented here. Richard Cooper was supported by Joint Council Initiative grant 8920199.

there is also a conspicuous absence of principles for combining these techniques. The manifestation of hybridness in extant hybrid systems is consequently somewhat *ad hoc*, with limited ramifications for general cognitive models. That the properties of connectionist and symbolic systems are in principle compatible (Hawthorn, 1989), and possibly critical to a cognitively plausible system (Clark, 1991), is clear. Important questions to be addressed concern, exactly how a hybrid system should be defined, and what kinds of constraint should be placed upon the design of such a system to ensure that it simultaneously manifests the virtues of both symbolic and connectionist systems.

1.1 Types of Constraints on Hybrid Systems

Possible constraints on hybrid systems fall into three classes. The first, *task dependent* constraints, flow directly from attempting to model a particular cognitive task in a hybrid system. Design choices are fostered by these constraints regarding both the symbolic and connectionist systems, and their interaction—choices which may not generalise to other kinds of task. The sole criterion for the validity of the system's particular manifestation of hybridness is whether it can model the task behaviour in question. Wermter & Lehnert's (1989) hybrid model of prepositional noun phrase interpretation is an example of a system employing this sort of constraint. Here, a symbolic system effects a restricted syntactic analysis, a distributed connectionist system acts as a semantic memory for noun/preposition relations, and a localist network integrates their outputs. Although a viable instance of a concrete hybrid system, it is not clear that the model can generalise to other types of noun phrases, or larger syntactic structures.

The second class of constraint, *system dependent* constraints, depend upon the designer's intended relationship between the connectionist and symbolic functioning. Beginning with either a symbolic or connectionist system, these constraints determine the ways in which the complementary type of system might emerge. In this way, the connectionist system constrains the nature of the symbolic, or vice versa. Such constraints do result in hybrid systems of a less distinctly *ad hoc* flavour than task dependent constraints, but it is clear that such systems might nonetheless be limited in their general cognitive applicability by the system chosen as starting point. That is, although available connectionist and symbolic systems are designed to have particular properties (e.g., content addressability and semantic perspicuity, respectively), a hybrid system derived from either type of system may be limited by the lack of transfer of many of these properties between types of system. This is exemplified by certain connectionist implementations of symbolic models: in Touretzky & Hinton's (1988) connectionist implementation of a production system, for example, it is not clear precisely how the implementation might generalise both to more complex production systems, and to different types of symbolic functions.

Task dependent and system dependent constraints are thus similar to the extent that the *justification* for particular constraints is somewhat weak: the ensuing constraint may apply to only a single instance of a hybrid system. In the task dependent case, a particular configuration of the symbolic and connectionist systems is designed to fulfill a single task only, so this configuration may not generalise to other kinds of task. In the system dependent case, such a configuration depends upon the precise design of the initial system, and hence different choices of initial system (which may, in some cases, be justified by particular tasks) lead to different hybrid system configurations. Again, then, particular configurations may not generalise across tasks.

These limited scope constraints contrast with *system independent* constraints: constraints that attempt to capture some fundamental desideratum on any cognitively plausible system. Such constraints emerge from quite general considerations regarding the functioning of cognitive agents, and may thus be defined independently both of any given task that the system performs, and of the realisation of the system. System independent constraints may include general properties of, or requirements on the performance of, all of the tasks an agent carries out. In respecting a system independent constraint, the precise form of both the connectionist and symbolic systems may be constrained, as may the relations between them. A consequence of the pre-theoretic origin of these constraints is that they may (indeed must) be defined independently both of any given task that a system performs and of the realisation of that system.

The distinction between system independent and task dependent constraints may be blurred in that the former might be said to be instances of the latter concerning a task of general application. The distinction may be sharpened via the faculty psychological distinction between horizontal and vertical faculties (Fodor, 1983). Task dependent constraints are constraints that may foster a design of a hybrid system *as if* the function it computes were indicative of a vertical cognitive faculty. Such faculties are domain specific, genetically determined, associated with distinct neural structures, and computationally autonomous (Fodor, 1983: 21). The design of the hybrid system is thus directed solely at computing the goal function, so compatibility with the computation of other functions is irrelevant. The architectural relations and the nature and computation of the algorithm may well be explicitly, deliberately dedicated, or they may simply *de facto* not generalise to other faculties. In contrast, system independent constraints operate across faculties in a horizontal manner: they inform functions across domains and therefore must be respected by cognitive faculties regardless of their architectural dedication and organisation (although, as Fodor (1983: 13) notes, the notion of a completely horizontal constraint—one that is respected by all mental faculties—may well be an idealisation). System independent constraints, as such horizontal constraints, thereby constrain a multitude of vertical cognitive faculties.

Although system independent constraints may be broader in *scope* than either task dependent or system dependent constraints, they are potentially weaker in *force*. That is, although system independent constraints are of general application across a spectrum of hybrid systems, the degree of detail determined by a particular constraint in a particular instance may not be great: a given system independent constraint might place only loose limitations on the configuration of an actual hybrid system. In contrast, task dependent and system dependent constraints, by definition, result in quite precise limitations on the manifestation of hybridness. This difference in scope and force may be attributed to the horizontal nature of system independent constraints. The scope of a constraint reflects the extent of its horizontality, whereas the force mirrors the amount of detail fostered regarding a vertical function. System independent constraints are thus necessarily of broad scope, but may have varying force depending upon the constraint in question.

System independent constraints may ensue from the requirement to map some aspect of the functioning apparent in one component of a hybrid system to a corresponding aspect in the complementary component. Preservation of varying degrees will express system independent constraints of varying force.

1.2 Interruptibility: a System-Independent Constraint

The ability to handle interruptions (henceforth referred to as “interruptibility"), is an aspect of functioning of considerable scope and importance.[1] Any cognitive system that exists in a world whose impinging events it cannot totally predict must be able to respond to events in its environment. These events, by their unpredictability, must constitute interruptions to ongoing behavioural sequences and cognitive processes. Such unpredictability includes both aspects of a constant world that cannot be anticipated (e.g., threats from predators) and “plastic" or “changeable" aspects of the world in which the rules according to which the environment operates change (Clark, 1991). Interruptions need not, however, be exogenous in cause. Particular *physiological* events or changes may have the effect of interrupting a cognitive system (for example, strokes, heart-attacks, and brain damage). Phylogenetically, the ability to handle interruptions will clearly confer species advantage; and ontogenetically, such an ability may be a necessary precursor to learning.[2]

Interruptibility, and in particular the preservation of interruption types between the component systems of a hybrid system, yields a system independent constraint of significant force in that it can provide for quite detailed constraints on the configuration of a hybrid system. In Section 2 we discuss the nature of interruptions, focusing on their categorisation and potential formal treatment. Our method of delineating the interruptibility constraint consists firstly of specifying interruption types in the component systems (Section 3 and Section 4) and then considering the possible correspondences between these types in various classes of hybrid systems (Section 5). Finally, Section 6 draws some general morals for the study of cognition.

2 The Nature of Interruptions

We take an interruption to be any disturbance to the normal functioning of a process in a system. Typically the cause of such disturbances is an unexpected communication event. Although a formalisation of an “unexpected communication event" would take us beyond the scope of this paper, we suggest that a promising avenue is to integrate insights regarding the conditional nature of information processing provided by Situation Theory (Barwise & Perry, 1983) with Milner’s (1989) transitional semantics for communicating concurrent systems.

Interruptions can be characterised according to a number of different parameters, yielding an interruption’s *profile*. Preservation or otherwise of profiles between components of a hybrid system gives rise to system independent constraints of varying strength (c.f., Section 5.2). Parameters within a profile include:

Source: Interruptions may have sources endogenous or exogenous to the system (or subsystem) interrupted. This distinction depends upon the system’s boundaries: an interruption exogenous to one system may be endogenous to an encompassing system.

Effects: Two parameters of the effects of a successful interruption (i.e., one which alters the system’s behaviour) are degree and extent. Degree concerns how much “damage" an interruption produces. In state transition terms (given a metric over states), degree may be formalised in terms of the divergence of an actual consequent state from the expected consequent state. Independent of this, an interruption’s extent may or may not be localised: given a characterisation of a system in terms of communicating subsystems, effects may be confined to relatively few or many of those subsystems.

Content: An interrupting signal may or may not have content over and above the fact that it is an interruption. We may differentiate between interruptions whose content does or does not influence the subsequent behaviour of the system. In state transition terms we may distinguish *consequent-state encoders* and *consequent-state non-encoders*.

Applicability: Certain signals may only have interrupting force if they impinge upon a system when it is in a particular state, rendering them *conditional*. A state transition interruption may thus “target" a particular antecedent state (or class of states), as well as encoding a particular consequent state.

Duration: An interruption may be characterised along three independent temporal dimensions. Firstly, the interrupting signal itself may be *temporary* or *enduring*. A temporary interruption is one that affects a process for just one “cycle", whereas an enduring interruption has an ongoing effect for subsequent “cycles" (c.f., Section 3.1 and Section 4.1). Secondly, the duration of an interruption’s local effects (i.e., its effects on the subsystem receiving the interrupting signal) may also be temporary or enduring: a subsystem may acclimatise to an enduring signal, or it may never recover from a temporary signal. Finally, non-local effects due to the fanning out of an interruption to other subsystems may also be enduring.

Mechanism for Recovery: An interruption may also be characterised in terms of the mechanisms employed to handle and recover from it. Such mechanisms may be explicit (and perhaps dedicated) or implicit. It is possible to have interruptions even when there is an explicit interruption handler. In such a case, the precise timing and location of any particular interrupting signal can still be unexpected. This aspect of an interruption is clearly *specification relative*: an abstract specification of a system may not include an interruption handling mechanism, whereas a more concrete specification may.

State Space of the Underlying System: The underlying system may have a discrete or continuous state space. For a continuous state space, a successful interruption may be identified by a discontinuity in the first derivative of, or kink in, the state-time graph. This identifying characteristic is not available in discrete systems, where an interruption might be identified by a failure of the transition from the current state to the expected next state, given the defining state transition function. (This is, again, specification relative).

3 Symbolic Systems

3.1 A Characterisation of Symbolic Systems

For concreteness, we assume that symbolic systems comprise minimally a set of autonomous subsystems (or “agents"), possibly operating in parallel, where each agent is capable of input and output (and hence communication) and is characterised in terms of states and serial transitions between those states. Given an agent’s input vector, $\vec{\imath}$, a state transition function, T, maps the current state of that agent to its subsequent state (so if the state of an agent at time t is represented by s_t, then s_{t+1} is given by $T(\vec{\imath}_t, s_t)$). An output function maps the agent’s state to its output vector. Thus the output of the agent at time t (i.e., when it is in state s_t) is $O(s_t)$. These assumptions define a general class of system, broadly inspired by the work of Milner (1989) on concurrent communicating processes and illustrated in Figure 1, where the ovals represent communicating agents, A and B, and communication channels between these and other agents (including, perhaps, the environment) are represented by double lines. The inputs and outputs of the agent A are explicitly labelled.

This characterisation of symbolic systems is essentially “local" in considering only state transitions within a single agent. A more “global" characterisation may be given by considering the state transition behaviour of a system of interacting agents. In a closed system, with no exogenous communications (possibly construing the environment as an agent within such a system), we can express the state of the system as a whole as a vector, $\vec{S}_t$, whose components are the states of the individual agents. Associated with such complex states, a state transition function, $\vec{T}$, may be defined from the state transition functions of the individual agents, such that $\vec{S}_{t+1} = \vec{T}(\vec{I}_t, \vec{S}_t)$. Here, $\vec{I}_t$ represents the inputs of all of the subsystems: there is a component of $\vec{I}_t$ for every communication channel within the system. Because of the closure of the system, the output vector is, in a normally functioning system, just the input vector (i.e., $\vec{I}_t = \vec{O}(\vec{S}_t)$).

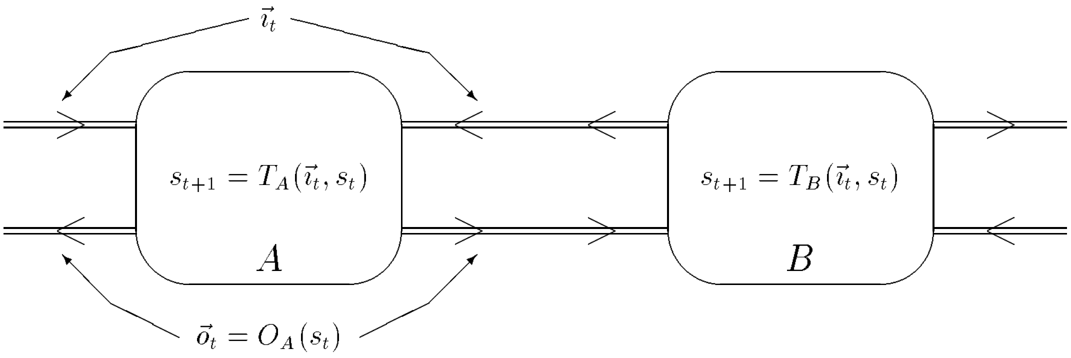
$\vec{\imath}_t$
$s_{t+1} = T_A(\vec{\imath}_t, s_t)$
A
$s_{t+1} = T_B(\vec{\imath}_t, s_t)$
B
$\vec{o}_t = O_A(s_t)$

Figure 1: A Prototypical Symbolic System

Hence, in a normally functioning system, $\vec{S}_{t+1} = \vec{T}(\vec{O}(\vec{S}_t), \vec{S}_t)$. This global characterisation allows a precise specification of types of interruptions in symbolic systems (Section 3.2).

The following examples illustrate the range of this class of symbolic systems:

SOAR: SOAR (Newell, 1990) can be viewed as a single agent working within the “cognitive band" and communicating with its environment via perceptual and motor agents. SOAR involves a working memory, a preference memory and a recognition memory, and its state may be defined by the collective contents of these memories. Processing in SOAR is cyclic, consisting of an *elaboration phase* followed by a *decision phase*, each of which can be construed as single state transitions. These phases constitute distinct processes, and a physical realisation may consist of a dedicated agent for each phase, with the elaboration agent communicating its results to the decision agent, which communicates its decision back to the elaboration agent, and so on. In this realisation of SOAR as two subagents, the elaboration agent would be required to pass long and complex messages to the decision agent. To minimise this communication an alternate realisation might also take each of the memories to be agents, such that an elaboration agent communicates mainly with the memories, only notifying the decision agent when a decision is required. The decision agent may then similarly communicate with the various memories, notifying the elaboration agent when a decision has been made. This realisation substantially reduces the traffic between the elaboration and decision agents, but at the expense of requiring further agents and significant communications between those and the original agents. In both of the above decompositions, the elaboration phase may be further analysed, revealing that it is itself cyclic, consisting of repeated *elaboration cycles*, each comprising an *input cycle*, a *preference phase*, a *working memory phase*, and an *output cycle*. Again, each of these processes may be the responsibility of dedicated agents. This decomposition reveals that, as a state transition, the elaboration phase consists of numerous intermediate transitions. The treatment of systems as communicating agents thus allows considerable flexibility in specification: one system may be decomposed in any of a number of different ways. Furthermore, each decomposition may have differing interruptibility characteristics. We discuss this further below.

M&M: Sloman (1987), in an account which we term M&M (“Motives and Mechanisms") suggests that a cognitive system must include mechanisms for dealing with motives—their generation, screening, comparison, scheduling, and satisfaction. Each of these processes may be the responsibility of a dedicated communicating agent (with perhaps several concurrent motive satisfying agents, corresponding to different effectors). We may thus phrase Sloman's motive processing in terms of a motive generating agent, a motive screening agent, a motive scheduling agent, a motive comparison agent, and possible many motive satisfying agents. The motive generator continually presents the screener with new motives. The screener transmits some of these motives on to the scheduler, blocking those which are not sufficiently insistent. The scheduler compares the motives it receives from the screener (via communications with the comparator), and passes one on to each satisfier. The mechanisms involved in this system are not sufficiently detailed by Sloman for us to provide a state transition characterisation for each of our agents. However, the input/output relations required of each agent suggest that with more detail a state transition specification would not be problematic.

3.2 Interruptions in Symbolic Systems

This type of symbolic architecture supports only two types of interruptions. *State transition interruptions* deny or over-ride an expected state transition *within* an agent. In contrast, *channel interruptions* interrupt the communication channel *between* agents, such that a signal sent from one agent is degraded in transmission or not received by the target agent. Generic instances of these types of interruptions illustrate many of the parameters of interruption profiles, as discussed below. Others, however, can only be properly addressed in fully-fledged systems. In Section 3.3, these parameters are considered in relation to the example systems.

State Transition Interruptions: The “global" characterisation of symbolic systems noted above allows a precise specification of state transition interruptions. Recall that, in a non-interrupted system, $\vec{S}_{t+1}$ = $\vec{T}(\vec{O}(\vec{S}_t), \vec{S}_t)$, or, eliminating the issue of the identity of the input and output vectors, $\vec{S}_{t+1}$ = $\vec{T}(\vec{I}_t, \vec{S}_t)$. If this equality does not hold—if the state at time $t+1$ is *not* given by $\vec{T}(\vec{I}_t, \vec{S}_t)$—then a state transition interruption must have occurred. On this picture, the state transition function must be read as only specifying an *expected*, rather than the actual, next state. Such interruptions are a local phenomenon: on the local level, a state transition interruption has occurred if s_{t+1} is not the output of $T(\vec{\imath}_t, s_t)$. If an interruption over-rides a state transition, then it must have content encoding the replacement consequent state; in state transition denial, however, the signal may be contentless, and may result in transition to a default state. In addition, the interruptor may encode content making its effects conditional upon some state of the system. Detailing the effects of state transition interruptions requires the global perspective on systems and a similarity metric for states. We may thus envisage state transition interruptions of small degree (where the distance between the pre-interrupted and the post-interrupted state vectors is small) and large extent (but where the number of differing components between the vectors is large, each component differing by only a small amount), or of large degree and small extent. In defining the duration of the interruption, we may take a cycle as a state transition. This, and the distinction between local and global characterisations of symbolic systems, grounds the duration of state transition interruptions: further analysis requires discussion of concrete systems.

Channel Interruptions: Channel interruptions may be characterised globally in terms of the identification of the input and output vectors. A channel interruption corresponds to a mismatch between these vectors ($\vec{I}_t \neq \vec{O}(\vec{S}_t)$), reflecting a breakdown of normal communication between agents. Channel interruptions *cannot* be characterised in local terms: they necessarily involve the communication channels *between* agents. We might associate a channel interruption with some distorting signal, whose source may be endogenous or exogenous, and the precise style of distortion fostered (ranging from delay, blocking and hastening, to filtering and transforming the content of communications) might be associated with the interruption's content. An interruption may be conditional if it is targeted upon particular sending and receiving subsystems which the channel connects (since channels themselves have no particular differentiating properties). Specification of the degree of the effects of a channel interruption requires a similarity metric over transmittable signals. This allows the degree of the interruption to be defined in terms of the divergence of the received signal from the transmitted signal. The extent of a channel interruption may be determined by the number of channels targeted by the interrupting signal in the sense discussed above, and a channel's cycle may be equated with one state transition in the sending subsystem, giving purchase on the duration of a channel interruption.

3.3 Examples of Interruptions in Symbolic Systems

SOAR: Newell (1990; 258–259) explicitly considers the possibility of interrupting SOAR. He notes two points: that the architecture's ability to propose any operator at any time means that any operation may be interrupted by some other operation; and the nature of the decision cycle means that any such interruption is the result of deliberation, and could thus have been ignored. The source of an interruption in SOAR may be the output of an elaboration cycle, or the environment (mediated by the input cycle). All such interruptions can only be assimilated in subsequent decision phases. Interruptions having their source in the environment demonstrate a form of specification relativity: in the third specification, whilst individual instances of input events are not expected (and are thus interruptions), the class of input events is explicitly catered for. It may further be noted that the specification in which each of SOAR's memories is regarded as a separate agent affords the possibility of interruptions to the channels between the memories and the other agents. These channel interruptions will result in incorrect reading of, or writing to, memories. In the specifications of SOAR which incorporate memories into agents, these “channel" interruptions will be manifest as state transition interruptions. Thus an interruption's type may be relative to a specification. This possibility is a consequence of the suggestion that agents may have internal structure consisting of communicating subagents. Clearly, in such cases, channel interruptions between subagents of an agent will normally result in state transition interruptions of the agent as a whole.

M&M: The scheduling of goals within Sloman's model implies that at any time a satisfier may be interrupted by the scheduler, so that a new goal may replace that satisfier's current goal and become active. Such satisfier interruptions are, within our specification, state transition interruptions. As the state transition nature of Sloman's agents has not been articulated, we cannot specify all of the details of such interruptions. We can, however, comment on the profile of permissible interruptions. The source of the interruption will be the scheduler. The interruption itself will be temporary, but its effects may be enduring (depending upon whether the interrupted goal is suspended or aborted). The mechanism for handling the interruption is explicit in the scheduler. Depending on the finer details of the specification, the interruption may or may not be contentless. If the specification involves something amounting to a blackboard for each satisfier on which the satisifier's current motive is written, then the interruption itself may be contentless, with the default next state being determined by the contents of the blackboard. Such a system would, of course, require the scheduler to write to the appropriate blackboard prior to interrupting its satisfier. Alternately, the scheduler might communicate the new motive directly to the relevant satisfier as the content of the interruption. The extent of the interruption will depend upon the proportion of the system's satisfiers required to satisfy the new motive. Assuming that that interruption stems from the acquisition of a high priority motive, if all the system's resources are required to satisfy that motive then the extent will be great. If, however, the satisfaction of the new motive only requires one of several satisifiers, then the extent will be limited.

Clearly, other forms of interruptions, interruptions which are in no sense anticipated, may be envisaged in SOAR and M&M if the possibility of damage or malfunction is entertained.

4 Connectionist Systems

A plethora of connectionist models has been advanced for a wide variety of tasks. These models are mostly based on what we term a *classical connectionist architecture*: we sketch such an architecture in Section 4.1 and in Section 4.2 discuss its interruptibility.

4.1 A Brief Characterisation of Classical Connectionist Systems

A classical connectionist architecture comprises a network of interconnected neuron-like nodes or cells, where each node can be viewed as an elementary calculation unit. The connections between nodes are (possibly) directed and activations pass between nodes along these connections. Each node thus receives activation from (possibly several) input connections, and (typically) if the total activation received by a node exceeds a threshold (which is specific to that node) then the node will fire, passing activation down its output(s). Connections are weighted, so that the activation received by a node from a single connection is not the output activation of its predecessor, but instead the product of this activation and the connection's weight. Given a certain activation, a unit's output is governed by output rules. Weights allow nodes to inhibit or excite the nodes to which they are connected: a large positive weight will tend to excite a subsequent node when its predecessor is excited, whereas a large negative weight will tend to inhibit the subsequent node. We entertain the possibility of "modular" networks comprising functionally discernible subnetworks. Where necessary in our discussion of interruptions, we distinguish between such systems interpreted both locally and globally.

The state of a connectionist system may be given in terms of the activations of its units. A system with n units may be described by an n-tuple of activations: the possible n-tuples defining the state space of the system. The activation of any unit, and hence the n-tuple, varies over time, tracing a state space trajectory. For systems continuous in time (whose movement from one state to the next is not governed by discrete time steps) with unbounded continuously differentiable (i.e., smoothly varying) inputs and outputs (such as a logistic function), this trajectory is smooth.

4.2 Interruptions in Classical Connectionist Systems

We identify two rudimentary sorts of interruptions in classical connectionist systems: *connection interruptions* and *node dysfunctions*. In connectionist systems, the *detection* of interruptions is dependent upon the continuity or otherwise of the system's trajectory in state space, and on the local/global interpretation of the system. In systems continuous in time with smooth input and output functions, and hence normally smooth state space trajectories, some interruptions can be detected as kinks in that trajectory. Any such kink *must* be attributed to an interruption, as they cannot arise given the ordinary functioning of any system. This holds for interruptions which are non-continuously differentiable or non-continuous. However, even in systems with (normally) smooth trajectories, some unexpected communications may be smooth. The manifestation of these smooth interruptions may be non-obvious in a system with a global semantics, but in a system with a local semantics we have an intuitively forceful characterisation of unexpectedness: excitations in semantically related units should be highly correlated, whereas in semantically unrelated units there should be no significant correlation. Violations of such correlations are unexpected. In virtue of this, the manifestation of interruptions in discontinuous systems with a global semantics (and some interruptions even in continuous systems with a global semantics) may be unclear. This is not to say that such interruptions do not occur, merely that under a global semantics their detection is problematic.

In considering the other parameters of an interruption in a classical connectionist system, we again discuss only those whose values can be specified for the bare classical connectionist case, and not others that require more detailed design choices than those entertained. In addition, it may be noted that no interruption can be prompted via the normal transmission of activation. Such transmission is the only endogenous means of communication, and so the only potential endogenous source of interruptions. All interruptions in classical connectionist (sub)systems must thereby be exogenous in source.

Connection Interruptions: In classical networks, transmission of activations between nodes is governed by weights on relevant connections. A connection interruption is an alteration of some connection's weight.[3] Connection interruptions may have content in encoding the details of the weight alteration they engender. Conversely, if connections have default responses to interruptions (such as weight zeroing), then such interruptions will be effectively contentless. In a globally interpreted system, since all connections are equal, there is no clear sense in which a connection interruption may be conditional. However, in a locally interpreted system, connections may be differentiated (in connecting nodes with distinct interpretations), and as such their interruption may be conditional. The degree of the effects of connection interruptions ranges from weight distortion (including random noise) to the zeroing or reversal of weights. Zeroing a weight corresponds to severing the output of the antecedent node, effectively removing the node from the system. Weight reversal corresponds to the inversion of excitation and inhibition. The extent of a connection interruption is the cardinality of the set of weights affected. Extent and degree of effects may also be conditional on the relationship between the current weights and the required results of the interruption.

Node Dysfunction: A network's nodes may also dysfunction, producing output at variance to the norm. This may be due to the miscalculation of a node's activation (perhaps due to an alteration of threshold), the miscalculation of the output from its activation, or the application of an incorrect output determining function. These possibilities may be indistinguishable in behaviour, however. Node dysfunction interruptions may have content in specifying facts about its effects on the target nodes, perhaps specifying a new threshold, or a replacement output function. The application of node dysfunctions may be conditional; for example, threshold changes are necessarily conditional in depending upon the difference between the prior threshold and the activation of the node. If a threshold change does not alter the sign of this difference, it will not alter the current output of the node. Further, in a locally interpreted system, node dysfunction interruptions may target particular nodes in virtue of their semantic interpretation.

4.3 Examples of Interruptions in Classical Connectionist Systems

Hinton & Shallice (1991) consider three types of “lesions" in their grapheme to sememe network in a simulation of acquired dyslexia. These are a proportion of the connections between each layer had their weights reduced to zero; random noise was added to the weights on these sets of connections; and in the two sets of hidden units a certain number of units were randomly excised. These three types of “lesion" are all subtypes of what we have labelled connection interruptions. The most direct means of achieving the excision of nodes is by reducing to zero the weights on the connections of that node (although the interruption might be cast in terms of node dysfunction, where the dysfunction maps all inputs to zero output); hence, the difference between Hinton & Shallice’s first and third types is one of location within the system, and not of type of interruption. The second type is a variant of the introduction of noise to the weights in a network, with the particular proviso that the unreliability of neural activity is simulated by the random value of noise selected for each “lesion".

In a mathematical characterisation of connectionist systems, Smolensky (1986) explicitly considers the effects of damaging particular nodes by setting their outputs to zero. As in the case of Hinton & Shallice, this corresponds in our characterisation to an interruption of the connections leaving the damaged nodes, although again the interruption might be cast as an appropriate node dysfunction. Smolensky’s analysis reveals that in a globalist interpretation of a connectionist system, excising a particular node of that system amounts to distributed damage at the semantic level. The techniques of linear algebra which he employs in discussing this form of interruption might equally well be applied to connection interruptions arising from other forms of weight modification.

5 Hybrid Systems

Applying interruptibility to hybrid systems, we first present a typology of hybrid systems, and then delineate some possible constraints on the relationship between interruptions across systems. Following this, we sketch the way in which examples of such hybrid systems conform to the interruptibility constraints.

5.1 Classes of Hybrid System

There is a primary distinction between *physically hybrid systems* and *non-physically hybrid systems*. In the former two distinct types of system (i.e., distributed connectionist and symbolic) are necessary components in the over-all functioning of the hybrid system: independently functioning symbolic and connectionist systems interact, and those functions are explanatorily necessary to an account of the operation of the system as a whole. A system composed of one or more connectionist systems serving one set of functions interacting with one or more symbolic systems serving another is physically hybrid (c.f., Wermter & Lehnert’s model noted earlier).[4]

Non-physically hybrid systems are those whose intended functionality is determined by one type of system (connectionist or symbolic), but whose performance of those functions is expressed in the terms of the complementary type. From a physical perspective, there is only one type of system, but hybridness is conferred by construing it, in abstract terms, *qua* the complementary system. Non-physically hybrid systems may or may not be constituted by some system dependent constraint flowing from one variety of system to the other. We term systems involving such a constraint *behaviourally hybrid systems*, and those not involving such a constraint *descriptively hybrid systems*.

In behaviourally hybrid systems, a given type of system (connectionist or symbolic) is taken as the starting point, and the behaviour of the complementary type of system (symbolic or connectionist respectively) is expressed in its terms. If a given set of functions is performable by one of the component systems, then a hybrid system derived from this system by some system dependent constraint will be able to perform all of the functions of this origin system, but will be able to do so in a way that encompasses some of the crucial properties of the complementary type of system. For example, a hybrid system whose origin system is symbolic will perform the functions of that symbolic system (in a connectionist technology), and will also accrue additional properties of the connectionist system (e.g., graceful degradation, generalisation, and so on). Conversely, a hybrid system whose origin is distributed connectionist will be able to perform the functions of the connectionist system, but will also have critical symbolic properties such as semantic perspicuity or symbolic explicitness. So, although there is only one actual, physical system, the properties and behaviours of the system are effectively hybrid. The critical point is that this hybridness arises from the clear operation of a system dependent constraint flowing from the origin system type to the complementary type, whereby the resulting hybrid system inherits certain of the properties of the complementary type.

In descriptively hybrid systems, the single physical system is not related to the complementary type of system by any system dependent constraint. Rather it can be interpreted or described as behaving, with some margin of error, after the fashion of the complementary type of system. As a result of this loose relationship, properties of the complementary system will not be acquired by the original system. Hence the attempt to interpret the behaviour of the hybrid system as concurring with the *modus operandi* of the complementary type of system will be open to error.

A further crucial parameter upon which these three types of hybrid system differ is concerned with the explicitness or otherwise of the symbols and rules that they employ. In particular, it is clear that a physically hybrid system will, in virtue of defining independent sets of functions, also employ both explicit symbols and explicit rules in its symbolic component. In contrast, those behaviourally hybrid systems for which the origin type of system is distributed connectionist, will have explicit symbols (in the sense, perhaps, of comprising a locally interpreted network), but only implicit (i.e., emergent) rules. And finally, a descriptively hybrid system will have neither explicit symbols nor explicit rules: the symbols and rules both being emergent.

5.2 Interruptibility Constraints

Hybrid systems entail the coexistence, at some level, of symbolic and connectionist components. Interruptions (of some discernible type) in one component may or may not correspond to interruptions (of some corresponding type) in some other component. Furthermore, aspects of interruptions, such as their profile, may be preserved to varying degrees. This mapping may operate from the connectionist to the symbolic or vice versa. In mapping, for example, from connectionist to symbolic systems, connection interruptions may or may not be mapped to interruptions in the symbolic system. In the case where interruptions *are* preserved, it still may be that particular types of interruptions are not preserved, in the sense that one type of interruption in the origin system may be mapped to different types of interruptions in the target system under different circumstances. (In the above example, some connection interruptions might be mapped to channel interruptions, and others may be mapped to state transition interruptions.) Thus, an even stronger constraint would be where the particular type of interruption *is* preserved in the mapping. That is, the types of interruptions are in one-to-one correspondence (in which case, all of the connection interruptions must either be mapped only to channel interruptions or state transition interruptions).

Hence, the three types of Interruptibility Constraint, in terms of decreasing specificity or force (defined with respect to a type of interruption in the origin) are:

Strong Coupling ($\mathbf{IC}_1$): For every interruption of a specified type in the origin system, there is an interruption of a unique corresponding type in the target system.

Loose Coupling ($\mathbf{IC}_2$): For every interruption of a specified type in the origin system, there is an interruption (not necessarily of a unique corresponding type) in the target system.

Open Coupling (IC$_3$): There may exist interruptions in the target system for which there is no corresponding interruption in the origin system.

Relative to a given interruption type, Strong Coupling is a constraint that demands an isomorphism between systems, whilst Loose Coupling demands only a one-to-many mapping. As a consequence of the relativity of this definition to a particular type of interruption, a hybrid system may be loosely coupled with respect to one type of interruption, but strongly coupled with respect to some other type. Open coupling provides for the case in which a specific connectionist interruption has generalised symbolic ramifications which typically may not ordinarily be construed as interruptions. This constrains to the extent that it requires a certain independence of the two components.

5.3 Interruptions in Behaviourally Hybrid Systems

Under the rubric of behaviourally hybrid systems, we take it that the system dependent constraint can flow either from the connectionist to the symbolic system, or vice versa. Examples of the former are systems employing the techniques of “relevance" and “skeletonisation" (Mozer & Smolensky, 1989); the latter subsumes the general field of symbolic systems that have been implemented in connectionist terms.

Mozer & Smolensky (1989) propose a technique for determining the relevance of each node to a task carried out by a connectionist system. They utilise this technique to produce a “skeletonised" network, in which less relevant nodes are “trimmed" away, leaving a smaller and more efficient network. In addition, they asseverate that the skeletonised network is semantically more transparent than the original distributed network: the skeletonised network is a smaller network with a localised semantics. To this extent the behaviour of the skeletonised network can be viewed as (explicitly) symbolic and (implicitly) rule-governed, in a way that is not directly open to the distributed underlying network. The only physical system is a connectionist one, but its behaviour has been trammelled so as to mimic symbolically driven behaviour.

In skeletonised systems, the system dependent constraint, the mapping between systems, causally flows “upwards" from the connectionist to the symbolic; hence the Interruptibility Constraint is similarly directed. The issue then arises as to which, if any, of the potential connectionist interruptions may occur in the skeletonised network. The first thing to note is that there is a generally applied weight zeroing (being themselves connection interruptions) in the very method of trimming the network of low-relevance nodes. Moreover, these interruptions are by definition without (significant) symbolic consequence. Indeed, Mozer & Smolensky claim that the symbolic behaviour — the performance of the functions — is enhanced. We thus have interruptions in the connectionist system which are not mirrored in the symbolic system (i.e., conforming to IC$_3$). However, for those nodes which are adjudged relevant, connection interruptions will necessarily be reflected in the symbolic system. This is simply because the skeletonised network is constructed only from the most relevant nodes of the origin system, so that any interruptions within the relevant nodes of the preskeletonised (origin) system *must* be reflected in the skeletonised (target) system.

A second type of behaviourally hybrid system is one in which the system dependent constraint flows “downwards" from the symbolic to the connectionist systems. The paradigm case of this is in connectionist implementations of symbolic systems. In order to determine precisely how interruptions to the symbolic system are correlated with interruptions in the connectionist implementation, we need to enquire into the precise nature of the implementation relation. Taking the line that this relation corresponds to the relation between Marr's Level 2 and Level 3 (Marr, 1982), an accurate implementation of a symbolic system must preserve both the representations of the symbolic system and the algorithmic orderings over them. If we take algorithms to be state sequences, then preserving algorithms entails the preservation of state sequences and hence of states. That is, in a correct implementation there must exist a direct mapping between states of the symbolic system and states of the connectionist system. Since we have defined symbolic systems in terms of communicating agents where each such agent is characterised in state transition terms, preserving states implies the preservation of agents. That is, agents of the origin/symbolic system must correspond to identifiable collocations of nodes within the connectionist/target implementation. This might be illustrated by considering an implementation of SOAR in terms of any of the specifications given in Section 3.1. Such an implementation will involve implementing each agent as a separable collocation of nodes, each performing the appropriate function (thus having, for example, a “decision" collocation and an “elaboration" collocation).

The implementation relation has direct consequences for the correlation between interruptions in symbolic and connectionist systems. In particular, since the relation preserves agents, interruptions to communications between agents (i.e., channel interruptions) must also be preserved as interruptions to connections (i.e., those connections acting as channels) between the connectionist implementations of those agents: there is an IC_1 constraint on channel interruptions. Further, since both states and algorithms are also reflected in connectionist terms (as particular patterns of activation, and as sequences of such patterns, respectively), state transition interruptions in the symbolic system cannot be tied to any one type of interruption at the connectionist level. For example, if a symbolic state is reflected in an activation pattern over a collocation of nodes, then an interruption to a state transition might be reflected in either a (set of) connection or node dysfunction interruption(s). Depending upon the implementation, there may be some way in which symbolic state transition interruptions do not have a reflection in the connectionist system; this is a possibility that we would wish to leave open. Hence there may be, at strongest, a loose coupling between state transition interruptions in the symbolic system and interruptions in the connectionist implementation.

In addition, the fact that the connectionist implementation is the only physical system in this type of hybrid model forces us to acknowledge the possibility that there may be interruptions to the connectionist level that have no clear reflection at the symbolic level. This is simply because *any* physical system is susceptible to interruptions. Consequently whilst IC_1 or IC_2 flow downwards, IC_3 flows upwards.

Note that the permutations of interruptibility constraints for the two types of behaviourally hybrid systems differ significantly. And furthermore the second, implementational strategy, manifests “differential couplings" for different types of interruptions.

5.4 Interruptions in Physically Hybrid Systems

Physically hybrid systems are the type that arise from “bolting together" symbolic and connectionist systems, each of which operate according to the standard patterns discussed earlier. In this kind of case, there are no clear system dependent constraints on their relationship. Rather, the pattern of relationships between interruptions in the two types of system is fixed by the way in which the systems are allowed to communicate. The physical existence of two types of communicating systems allows for complex interactions between the systems (that is, interactions both within a set of connectionist and/or symbolic systems, and between them): an interruption to the symbolic system may originate in the connectionist system, or vice versa. Such interactions may conform, in the most extreme case, to IC_1, but in general this need not be so. On the other hand, the least constrained possibility must allow for the interruptions in one type of system to have no answering interruptions in the complementary type. That is, it must allow for patterns conforming to IC_3, the precise manifestation possibly being dependent on the modularity of the component systems.

Given the variety of ways in which a physically hybrid system could be manifest, little concrete can be said without considering explicit examples. However, general morals can be drawn by reconsidering our characterisation of symbolic systems as communicating agents. A physically hybrid system will typically be comprised of such agents, where at least one is of a different type from the rest. For example, in M&M, the motive screener might be realised in connectionist terms, with the other agents being symbolic. There may thus be a causal flow of interruptions between agents, and hence between the symbolic and connectionist systems. Assuming open coupling, an asymmetry can be isolated concerning the effects of interruptions in such a system. Due to the graceful degradation of connectionist systems, interruptions with significant effect in the symbolic system may have less significant effects in the connectionist system to which they flow. Graceful degradation may thus minimise the effects of such an interruption in a physically hybrid system. Conversely, interruptions of minor effect in the connectionist system will generally have significant effects in the symbolic system to which they flow, simply because of the “brittleness" inherent in rule-governed systems.

Causal connections between the types of systems within a hybrid system (which is, by definition, the case in physically hybrid systems) leads to the possibility of a *causal* manifestation of Interruptibility Constraints. This refines our current constraint, which is strictly *correlational*, deliberately subsuming causal and non-causal instances. Thus, whilst we must still allow for the three strengths of Interruptibility Constraint detailed above (where an interruption of some type in the origin system may or may not cause an interruption of a particular type in the target system), there is further the possibility that the cause of an interruption (in the target system) may lie in the uninterrupted functioning of the origin system. This is conceivable in any system, most especially those which contain an explicit mechanism for coping with interruptions (since in such systems classes of interruptions, though not particular instances, are explicitly anticipated). These causal cases correspond to the reverse of IC_3: an interruption in the target system has no corresponding (causal) interruption in the origin system.

A concrete example of such interruptions might be provided by an alternate physically hybrid construal of M&M, where a connectionist scheduler interacts with symbolic satisfiers. In such a system the normal functioning of the scheduler will explicitly involve the interruption of satisfiers when a motive of higher than current priority is detected.

5.5 Interruptions in Descriptively Hybrid Systems

The label “descriptively hybrid system" might be thought of as something of a misnomer. The extent of hybridness embodied in such systems is the weakest of the three types that we have isolated: rather than one type of system behaviourally approximating the behaviour of the complementary type, it is simply that, from some particular perspective, the system can be *described as if* it were behaving in that way. Symbolic systems are often said to “approximate" the physical connectionist system (Smolensky, 1988; McClelland & Rumelhart, 1986). This necessary interpretive step motivates a slight scepticism about including such systems under the rubric of truly hybrid systems.

The correspondence between interruptions within descriptively hybrid systems depends on the descriptive adequacy of the symbolic approximation. If, for example, the symbolic approximation does not cater for exceptions to its rules, exceptions which *are* achieved via interruptions within the connectionist system, then it is clear that we have a case of IC_3. This kind of picture provides a way of construing the connectionist “lesions" which underpin Hinton & Shallice’s model of “acquired dyslexia". In this case, there is a set of precise interruptions to the connectionist level, which have a generalised effect on the symbolic system that might describe the grapheme-sememe relation. Conversely, it may be the case that exceptions to rules, which are not a consequence of interruptions in the connectionist system, are explicitly catered for via symbolic interruptions. In both cases interruptions in one system are not paralleled by interruptions in the other.

6 Some Consequences for the Study of Cognition

In this section, we sketch some consequences of the Interruptibility Constraint and of the general framework. First, we note various uses of the Interruptibility Constraint as different forms of constraint on hybrid systems. Second, we consider the implication that different kinds of constraints between connectionist and symbolic systems might hold in a single hybrid system, for different cognitive functions or processes (or sub-processes). Lastly, we adumbrate some other candidate system independent constraints.

How do system independent constraints constrain hybrid models? The framework we have developed is strictly *correlational*, in that it considers the coupling of aspects of functioning across components of a hybrid system. There are, however, two ways in which this correlational descriptive device may be manifest as a constraint in the construction of cognitive models. Firstly, we can employ the potential couplings as *causal* constraints as in the operation of physically hybrid systems (as noted above). Secondly, we can employ them as constraints on the *design* of hybrid systems. Such a use can only hold for physically hybrid systems and behaviourally hybrid systems. One design use is to employ an Interruptibility Constraint as a system dependent constraint in the following way. The design of a connectionist system might be conducted so as to allow for the kinds of interruptions that we have defined; then we might employ IC_1 as a constraint on the set of permissible symbolic systems: any acceptable symbolic system must relate to the connectionist system so as to provide for a strong coupling of interruptions. Exactly the same kind of constraint could flow “downwards" from a symbolic system to the set of permissible connectionist systems. A stronger possibility for the use of interruptibility as a constraint on physically hybrid systems is to take the initial design of *both* the connectionist and the symbolic systems as constrained by the need to provide for interruptions, and then employ IC_1 to constrain the relations between them in a hybrid system.

A stronger design use stemming from the requirements of interruptibility would be to motivate a reconsideration of the basic capacities of the component systems. Consider the case of connectionist systems. In classical connectionist systems, no input is differentiated from any other: the output is a function of the total input. One can envisage, however, more complex nodes which fire if and only if each of their inputs individually exceeds some threshold, or nodes which may only fire if some distinguished input exceeds some threshold, but whose behaviour is still dependent on remaining inputs (i.e., nodes which may be switched on or off by some distinguished input). Now, suppose that interrupting signals constitute a kind of differentiated input. An input-differentiating system can support a broader class of interruptions due to its finer-grained orientation on the effects of inputs on the generation of outputs.

These systems can support connection interruptions and node dysfunctions, but further allow for dedicated “interruption" connections, whose activation denies the “normal" firing of the node, even where the rest of the combined inputs would exceed the classical threshold. This kind of interruption can occur when there is no alteration to either the weights on the other connections of the functioning of the other nodes in the system.

Concerning cognitive modelling, the decisions as to which use to put the constraint to (i.e., causal or design), which form of coupling to employ, and which type of hybrid system to employ, are independent questions which we will not address here (c.f., Cooper & Franks, 1992a).

We noted in Section 5.2 that the definition of the variants of IC was relative to a particular type of source interruption; this implies that different strengths of constraint might hold for different kinds of source interruption. So there may be different couplings between the systems for different kinds of interruption. One way of generalising this is to *any* kind of system independent constraint (see below); another is to generalise it to a spectrum of cognitive functions and processes. This would result in a hybrid system in which the connectionist and symbolic systems betrayed different degrees of coupling relative to different tasks or functions. That is, the widespread assumption that there is a single relation between the connectionist and symbolic systems in a cognitively plausible hybrid system appears, from this vantage point, to underestimate the potential complexity in that relation. This potential for *simultaneously variable multi-dimensional couplings* (SVMC) between connectionist and symbolic systems provides for an intuitively powerful way of viewing a claim made often in the connectionist literature. This is that the traditional disjunction between “higher" cognitive functions (e.g., reasoning: usually modelled in symbolic terms) and “lower" cognitive functions (e.g., perception: usually modelled in connectionist terms) is better seen as a continuum (e.g., McClelland & Rumelhart, 1986). Our view is that the appearance of a continuum is fostered by there being many different degrees of, and dimensions for, the coupling between symbolic and connectionist systems, each perhaps relative to either a given set of functions, or even to particular functions. In brief, different such functions may foster “differential couplings" between the systems deriving from overall SVMC. The appearance of a “grand continuum" of all cognitive processes, then, issues from the coarseness of the perspective from which the functions are typically viewed.

A final question concerns the specification of further types of system independent constraints. Recall that a system independent constraint is a type of constraint that flows from general cognitive considerations, which can be defined in a system independent manner, and which should be reflected in the operation of a fully-fledged hybrid system. The possibilities that we suggest here are avenues for further research. Such research may, we suggest, indicate that the various constraints are not strictly orthogonal.

We note three cases. Firstly, there is the case of representation. Systems typically “internalise" their symbols. That is, symbols implicated in processing have analogues within the processor. This is often the case in both connectionist (recall the issue of the interpretation of nodes) and symbolic systems, but is by no means necessary. Turing machines, for example, may have no internal representation of the symbols over which they compute, in that the “representation" is restricted to an external tape. Clark (1989) has also suggested that this “externalisation" may be employed in connectionist systems. Hence the capacity to manipulate both internal and external symbols is a potential dimension for a system independent constraint on hybrid systems.

Another possibility is that the relationship between the connectionist and the symbolic within a hybrid system may preserve processes: processes within the symbolic system may correspond to identifiable processes within the connectionist system. To elaborate, a process in the symbolic system might be identified as a sequence of state transitions with an attendant semantics; and, within a connectionist system, a sequence of patterns of activation with its semantics may be viewed as a process. To the extent that there is a well-defined mapping between these two semantic interpretations, we should therefore moot degrees of congruence between the associated state transitions and patterns of activation. Furthermore, it may be the case that this mapping holds only above a particular level of granularity. That is, the systems may exhibit corresponding processes without necessarily exhibiting corresponding subprocesses.

A final possibility concerns the claims to modular organisation of the cognitive system (Fodor, 1983). In essence, for those faculties designated as modular, we might expect the "vertical" functional cleavages to be preserved across components of a hybrid model. That is, where there is a distinct symbolic module for a function, this should be mirrored by a connectionist module. Note that although the faculties themselves are vertical, the modularity constraint operates in a horizontal manner, across modules: the same qualities define a module regardless of its dedicated task. Modularity can thus be posited as a genuine system independent constraint.

The potential non-independence of system independent constraints is well illustrated by the interaction of modularity and interruptibility. These constraints are, *prima facie*, in conflict. For example, interruptibility suggests that processing need not be mandatory, and that modules need not be informationally encapsulated (in cases where an interruption's source is exogenous to the interrupted module). The resolution of such conflicts would seem to require some prioritising of constraints (Cooper & Franks, 1992b).

7 Conclusions

We have attempted to establish the plausibility of developing a novel type of constraint on hybrid systems: system independent constraints. Our vehicle for this has been a detailed examination of interruptibility, as it applies to symbolic, connectionist, and hybrid models. System independent constraints were motivated from within a general framework in which issues concerning hybrid modelling of cognition can be discussed. This framework involves a classification of hybrid systems, and a typology of constraints which can be put to various uses in modelling. The utility of this framework has been demonstrated in several ways. First of all, the definitions of the component systems of a hybrid model apply naturally to the explication of interruptions in extant symbolic and connectionist models. Secondly, the implication of SVMC in cognitive functioning provides for a sophisticated interplay between the components of a hybrid model. Thirdly, the potential interactions between candidate system independent constraints suggests an interesting avenue of debate and research which can augment the empirical evaluation of such constraints. Clearly, the plausibility of the particular constraint considered is independent from the viability of the framework. Although the application to interruptibility demonstrates the utility of the framework, its general vindication and assessment will depend upon applications in broader contexts.

References

Barwise, J. & Perry, J. (1983), *Situations and Attitudes*, Cambridge, Mass.: MIT Press.

Clark, A. (1989), *Microcognition*, Cambridge, Mass.: MIT Press.

Clark, A. (1991), 'In defense of explicit rules', in W. Ramsey, S. P. Stitch & D. E. Rumelhart, eds., *Philosophy and Connectionist Theory*, Hillsdale, New Jersey: Lawrence Erlbaum, pp. 115–128.

Cooper, R. & Franks, B. (1991), 'Interruptibility: A new constraint on hybrid systems', *Artificial Intelligence and the Simulation of Behaviour Quarterly,* **78**, pp. 25–30.

Cooper, R. & Franks, B. (1992a), 'Types of Constraints on Hybrid Models of Cognition'. In submission.

Cooper, R. & Franks, B. (1992b), 'Modularity and Interruptibility: On the Interaction of Constraints on Models of Cognition'. In preparation.

Fodor, J. A. (1983), *Modularity of Mind*, Cambridge, Mass.: MIT Press.

Hawthorn, J. (1989), 'On the compatibility of connectionist and classical models', *Philosophical Psychology,* **2** (1), pp. 5–15.

Hinton, G. E. & Shallice, T. (1991), 'Lesioning a connectionist network: investigations of acquired dyslexia', *Psychological Review,* **98** (1), pp. 74–95.

Marr (1982), *Vision*, San Francisco: Freeman.

McClelland, J. & Rumelhart, D. (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2,* Cambridge, Mass.: MIT Press.

Milner, R. (1989), *Communication and Concurrency,* London: Prentice Hall.

Mozer, M. C. & Smolensky, P. (1989), 'Using relevance to reduce network size automatically', *Connection Science,* **1** (1), pp. 3–16.

Newell (1990), *Unified Theories of Cognition*, Cambridge, Mass.: Harvard University Press.

Rumelhart, D. & McClelland, J. (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1,* Cambridge, Mass.: MIT Press.

Sloman, A. (1987), 'Motives, mechanisms and emotions', *Cognition and Emotion,* **1** (3), pp. 217–233.

Smolensky, P. (1986), 'Neural and conceptual interpretation of PDP models', in McClelland & Rumelhart (1986), pp. 390–431.

Smolensky, P. (1988), 'On the proper treatment of connectionism', *Behavioural and Brain Sciences,* **11**, pp. 1–74.

Touretzky, D. S. & Hinton, G. E. (1988), 'A distributed connectionist production system', *Cognitive Science,* **12**, pp. 423–466.

Wermter, S. & Lehnert, W. G. (1989), 'A hybrid symbolic/connectionist model for noun phrase understanding', *Connection Science,* **1** (3), pp. 225–272.

Notes

[1] Although there is an intriguing relationship between our notion of interruptibility, and the discussion of interrupts in computer science, this issue can fend for itself for the present. Our present goal concerns only interruptibility within cognitive systems.

[2] In the terminology that we develop in this paper, learning can be seen as the development of an explicit mechanism to cater for a particular class of interruptions.

[3] We here assume that “normal" functioning of a network involves fixed weights. We are not concerned with systems in their learning phase.

[4] It should be clear that the class of systems covered by the label *physically* hybrid systems includes abstract systems whose functional definition *necessitates* the specification of both a connectionist and a symbolic system.