

ОТ КРИТИКИ КОННЕКЦИОНИЗМА К ГИБРИДНЫМ СИСТЕМАМ ОБРАБОТКИ ИНФОРМАЦИИ

*Цепцов В.А.,
Институт психологии РАН.¹⁴*

Коннекционизм, заметные успехи которого были с некоторой долей непонимания отмечены Фодором и Пылишиным во введении к недавно (1995) переведенной на русский язык статье, остался практически незамеченным в отечественной психологии. Это было обусловлено, с одной стороны, мощными теоретическими традициями психологии деятельности, общения и нейролингвистики, которые создавали консервативный фон по отношению к нововведениям, с другой - катастрофическим отставанием компьютерных наук и соответственно почти полным отсутствием междисциплинарных исследований в этой области.

В отечественной психологии время от времени появлялась умозрительная критика кибернетического подхода, кибернетической метафоры, однако это относилось в большинстве случаев к общим вопросам сопоставления "живой" психологии и "мертвой" машины. В западной психологии уже началась эпоха когнитивной психологии, которая ставила своей задачей изучение процессов обработки информации человеком и для нее, имеющей перед глазами достижения разработчиков искусственного интеллекта, эта разница не было столь уж очевидной - достоинства этой метафоры были рассмотрены Б.М. Величковским (1982, сс. 54-66).

Действительно, когнитивная система человека обрабатывает информацию, и так называемая "компьютерная метафора" в отношении психологической системы не так уж далека от реальности. Еще со времен Декарта, "Хомо Механикус" стал одной из полноправных ипостасей человека в физическом мире. Согласно тезису о различении протяженной смертной субстанции (кости, мускулы, нервы и т.д.) и бессмертной непротяженной (мысли), человек в "протяженном" измерении целиком сопоставим с ма-

¹⁴*Работы выполнены при финансовом содействии Российского Фонда
Фундаментальных Исследований N96-06-803496 и Российского
Гуманитарного Научного Фонда N95-06-17491

шиной. Посылка, возникшая на фоне средневекового уровня развития техники, в эпоху компьютеров приобретает совершенно новые черты.

Что представляет собой собирательный антропоморфический образ компьютера сегодня?

Даже у бытового современного компьютера есть органы восприятия развитые достаточно для того, чтобы различать образы и символы с более высоким разрешением, чем у человеческих органов чувств: сканеры, цифровые видеокамеры обеспечивают зрительное восприятие, микрофон - ввод звука, клавиатура - аналог системы тактильных рецепторов. У компьютера есть система декодирования аналоговых сигналов (перцептивных образов, получаемых от устройств ввода) в универсальный внутренний язык электрических импульсов, бегущих по кремниевом-металлическим цепям, который в свою очередь преобразуется в "ментальный" язык кодов центрального процессора, доступных "пониманию" на вышележащих уровнях виртуальной машины.

Об этой виртуальной машине следует сказать особо. Своим отделением от компьютерного "железа" она обязана в первую очередь работам Ньюела, Саймона (1972) и Пылишина (1984). В значительной мере она является аналогом знаний человека и представляет собой совокупность элементарных субстанциональных единиц и набор операций-примитивов, которые могут быть к ним приложимы. Примеры виртуальной машины - это интерпретаторы языка АДА, Си, ЛИСП, Пролога или любого другого языка программирования. Компьютер, благодаря виртуальной машине, понимает все правильно построенные выражения на том или ином языке, а в расширенном виде выдает сообщения об ошибках, делает подсказки для их исправления. Виртуальная машина выступает посредником-переводчиком между человеком, который формулирует свои запросы и команды на одном из языков высокого уровня (примеры выше), языком центрального процессора (языком низкого уровня), который в свою очередь передает команды периферическим "органам": обычно это дисплей - устройство перевода закодированной информации в изображение. Принтер или плоттер служат для письма и рисования. Благодаря устройствам обработки звука обеспечивается как вывод образцов речи, так и ее синтез.

Итак, современный компьютер представляет собой создание достаточно высокой сложности и заслуживает сравнения с чело-

веческой когнитивной системой, что вынуждает психологию по-новому рассматривать этого постоянно "эволюционирующего" конкурента человеческой психики. Начиная с 80-х годов стали все чаще появляться работы, в которых компьютер рассматривается как совершенная модель человеческой когнитивной системы, но никак не макроструктурный аналог (Rumelhart, 1979); Abdi, 1993). Модель кибернетическая, которая состояла из функциональных блоков ввода, хранения и переработки символической информации, центральное место которой занимала машина Тьюринга, приводящая систему в состояние соответствующее символу на входе, была подвергнута критике. Эта критика была направлена на тех исследователей, которые представляли психологическую систему обработки информации в качестве линейной структуры "ленточного" типа. Такими моделями изобилуют работы по обработке речи, текста, принятию решений. На наш взгляд, обвинение того или иного автора в том, что из-за его модели проглядывает компьютер, лишь на том основании, что автор представляет переработку сигнала линейно или использует понятия программирования, несостоятельно, так как онтологически это неверно. Скорее разработчиков компьютеров следует обвинить в психологической метафоре при конструировании блок-схем для обработки информации внутри своих изделий, ведь очевидно, что вся внешне наблюдаемая психологическая активность человека связанная с обработкой символов (речь, чтение, письмо) организована линейно.

Критика работ, которые рассматривали психологические механизмы независимо от того, метафорой каких механических или электронных устройств они являлись, была методологически верно и необыкновенно просто построена коннекционистами. Сначала был сформулирован вопрос о том, какая метафора может быть более точной в отношении психики. И вполне естественно, что ответ был: "мозг". То, что долгое время было источником психофизиологической проблемы и разделяло материальный субстрат с психикой, стало в работах коннекционистов объединяющим моментом.

Факты: мозг состоит из нейронов (около 10^{11}), нейроны объединены между собой входными (дендритами) и выходными (аксонами) связями, каждый нейрон принимает сигналы и передает их, нейроны организованы послойно. Каждый нейрон работает параллельно остальным, что делает его качественно отличным

интегратором информации по сравнению с элементами линейно организованных схем. Все это, как и другие многочисленные свидетельства анатомического строения и физиологических функций мозга, делает его незаменимым устройством, на фоне которого появляется психика и далее вся сложная система психологических проявлений его функционирования. На основе представления нейрона в качестве элементарного прототипа единицы "когнитивной" структуры в коннекционизме был сформулирован главный тезис о параллельной распределенной обработке информации (parallel distributed processing - PDP, Rumelhart and McClelland, 1986).

В работах коннекционистов элемент сети нередко идентифицируется с нейроном, как это делают представители сформировавшихся в отдельное направление нейронаук. Порой это приводит к смешению подходов, однако коннекционизм это прежде всего абстрактная модель, которая вводит абстрактное понятие единицы (unit), называемой иногда клеткой, иногда ячейкой, иногда процессором, и связи (connection), которая представляет собой информационный канал ограниченной пропускной способности.

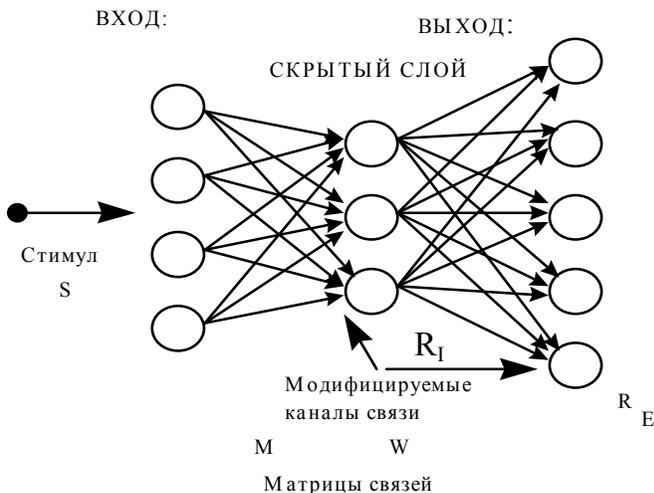
Важным достоинством коннекционизма стало плодотворное взаимодействие психологов и разработчиков систем искусственного интеллекта, что позволило осуществлять быстрый переход от теоретических моделей к виртуальным сетям, которые моделировались на компьютерах, и эмпирическим данным, получаемым на основе применения коннекционистских алгоритмов в психологическом эксперименте. Эту схему легко увидеть в работе по верификации модели ШОРТЛИСТ (SHORTLIST), которая позволила сравнить теоретические выкладки относительно процесса распознавания слов в устном высказывании: 1) посредством соревнования множественно активированных лексических гипотез и 2) чувствительностью к просодической структуре предложения (McQueen, Norris, Cutler, 1994).

Первая материальная реализация сетевой модели восприятия - Перцептрон была создана всего четыре десятилетия назад и названа создателем Марк 1 (Rosenblatt, 1961). Она состояла из 400 элементов (20 X 20) и занимала пространство почти в две комнаты. Современный серийный сканер - аналог воспринимающей изображение сетчатки - имеет физическое разрешение 800 X 800 точек на дюйм, то есть, на площади приблизительно в 6.25 кв.см.

размещено 640000 воспринимающих элементов, выполняющих работу, похожую на перцептрон Розенблатт.

Элементарная коннекционистская модель (Рис.1) состоит из конечного числа элементарных процессорных единиц (входной слой из I элементов, выходной слой из J элементов), которые объединены связями с изменяемой за счет того, что каждый элемент может изменять пропускную способность канала, интенсивностью, называемую обычно весом. Связи элементов представляются в виде матриц весов: M и W . Каждый входной элемент связан с каждым выходным элементом: $M = K \times I$, $W = J \times K$. Архитектура наиболее простой сети включает только два слоя - входной и выходной, однако такая сеть не может, например, обучаться нелинейно. Введение промежуточного скрытого слоя (K) позволяет устранить этот недостаток. Любой стимул в этой модели соотнесен с вектором активации входных элементов сети (I), множество связей объединено в матрицу, и ответу сети также соответствует вектор активации элементов (J).

Рис.1. Архитектура сети с промежуточным скрытым слоем



Цель коннекционистской сети (это понятие очень важно, так как позволяет уже на этом этапе ввести не функциональную, а интенциональную трактовку поведения живой системы) в том, чтобы ассоциировать вектор входа с вектором выхода или трансформировать стимул в ответ. Простым примером ассоциативной работы будет установление конфигурации параметров связей (заполнение матрицы, которая, условно, была пуста, значениями весов), другими словами, между элементами сетевой структуры устанавливается ассоциативная связь, выраженная матрицей весов, а в матрице количество строк соответствует множеству элементов первого слоя (входного), и количество колонок - множеству элементов второго слоя (выходного).

Основная задача сети, подобной перцептронну, - обучиться правильно реагировать на стимул, то есть поставить вектор входа в соответствие с вектором выхода. Для этого сеть должна "видеть" или "знать" ответ, который необходимо вызвать, а также получать информацию об ошибках, из чего следует, что эта сеть предполагает наличие вышележащего управляющего уровня. Обучение осуществляется каждым "нейроном" независимо (правило локального научения), и он меняет пропускную способность своего "синапса", если дает ошибочный ответ. Если "нейрон" (в абстрактном представлении - активный элемент) на выходе не активирован в тот момент, когда должен быть активирован, он увеличивает пропускную способность "синапса" (входного канала элемента). Основные проблемы, которые рассматриваются при изучении обучения сетей, - поиск алгоритмов обучения за минимальное число проб, поиск архитектур с минимальной сложностью. Среди правил научения можно отметить правило Видроу-Хоффа (Widrow-Hoff, 1960) или правило Дельта (Rumelhart, McClelland, 1986).

Основная отличительная черта коннекционистских моделей обучения в том, что в них процесс обучения управляется данными (стимулами), то есть изменение стимула влечет за собой переучивание сети. В классической когнитивной психологии научение происходит с привлечением знаний, заложенных в символической форме, например, ментальных репрезентаций или концептов, и на основе формализмов, которые нередко имеют вид общих правил. То есть, например, следуя логике научения с позиций теории деятельности, процесс операционализации идет от высших "сознательных" форм действий к свернутым операциям.

На более сложных коннекционистских моделях, которые позволяют учитывать нелинейные эффекты, изучается процесс обратного сообщения (back-propagation, см. подробнее Rumelhart, Hinton, Williams, 1986). Для таких моделей характерно наличие в промежутке между слоем элементов входа и слоем элементов выхода одного или более промежуточных слоев (на Рис.1. слой элементов К), которые исполняют роль модификаторов связей. Смысл обратного сообщения в том, что при вычислении ошибки между желаемым ответом и ответом, полученным в элементе выходного слоя, меняется пропускная способность связи в элементе выходного слоя, и одновременно возвращается значение сделанного изменения в обратном виде на элементы промежуточного слоя, которые в свою очередь меняют пропускную способность собственных связей, что позволяет изменять вес связей нелинейно.

В целом работы коннекционистов сложны для восприятия даже тех психологов, которые легко ориентируются в статистике традиционной когнитивной психологии. Эта сложность - следствие, на наш взгляд, того, что сама плоскость сечения психологической системы сдвинута настолько, что человек, понимающий вычислительную природу семантического пространства, структуры ментальных репрезентаций или конструкторов сознания, не может перенести эту плоскость настолько, чтобы узлы семантической сети стали элементами подобными коннекционистским элементам. Эта особенность коннекционизма и вызвала критику Фодора и Пылишина (1995), на которой мы не будем останавливаться, отсылая читателя к статье, отметим лишь, что статья вызвала широкий отклик и породила немало методологических дискуссий, в ходе которых аргументы авторов понемногу теряли свою изначальную стройность и вес.

Чалмерс (Chalmers, 1993) подверг критике ряд положений вышеназванной статьи, доказав, что эмпирические данные, полученные исследователями коннекционистского подхода, обладают системностью. Большой объем замечаний Фодора и Пылишина не слишком хорошо соотносится с теми аспектами коннекционизма, которые они выбрали в качестве мишени. Как будто сторонники коннекционизма утверждают, что если кирпичики сложить в кирпичики, кирпичики сложить в стены, а стены накрыть крышей, то получится здание, в то время как сторонники традиционного подхода ставят им в упрек то, что из этого все равно не

получится, к примеру, Нотр Дам. Ограничение коннекционистского подхода рамками локальной сети и анализ его вне условий развития облегчили авторам задачу, однако, как отмечает Смоленский (Smolensky, 1988), эта критика отражает не слишком верный взгляд на коннекционизм. По его мнению, субсимволическая парадигма коннекционизма позволяет ему занять важное место в формировании когнитивной психологии, которая будет иметь прочную методологическую связку с науками о мозге. Анализ когнитивных процессов на субсимволическом и субконцептуальном уровнях позволяет рассматривать когнитивную систему как вычислительную систему. Конечно, позиция Смоленского выглядит настолько твердой, что порой в ней можно видеть попытку трактовать коннекционизм как альтернативу существующей когнитивной психологии. Маринов (Marinov, 1993), в своей статье оспаривая это "нападение", отметил, что проведенное им сравнение эффективности научения искусственных нейронных сетей, использующих алгоритм возвратного сообщения, с алгоритмами деревьев индуктивного вывода не привели к появлению результатов, позволяющих заявлять о преимуществе какого-то из подходов.

Статья Фодора и Пылишина, вызвавшая множество откликов сразу после своего появления, периодически вспоминается коннекционистами по настоящее время, так как новые разработки и научная рефлексия ведут исследователей к усовершенствованию моделей.

Работа Бечтела (Bechtel, 1993) отражает тот прогресс, который происходит в коннекционизме вслед за критикой извне и в результате естественного развития самого направления. Во-первых, он выступает за введение модульности в коннекционистскую систему обработки информации, при которой несколько сетей осуществляют одновременный анализ задачи, а сеть регулирующая отбирает ту, решение которой наиболее эффективно. Во-вторых, дальнейшее развитие возможностей сетевого представления систем обработки информации связано с использованием рекуррентных входов, которые делают возможным доступ к информации от первых этапов обработки на более поздних этапах. В третьих, при научении в коннекционистской сети необходимо допускать не только изменение весов, но и уровня активации единиц, что даст возможность снижать размерность сети.

Еще более развернутую программную работу по преодолению недостатков коннекционистского подхода, выявившихся в ходе исследований последнего десятилетия, провели Рой, Говайл и Миранда (1995), которые предлагают обновить принципы научения в коннекционистских нейронных моделях в противовес классическим моделям коннекционистского научения. По мнению авторов, прежняя теория предполагает наличие сетевой организации мозга *ad hoc*, в ней допускаются законы локального научения и научение без опоры на память (примеры не сохраняются в памяти для последующего научения), а эти допущения не соответствуют особенностям мозга. Они сформулировали четыре пункта, которые вместили все наиболее важные, с точки зрения дальнейшего развития, особенности научения в сетях:

- задача оптимальной организации сети: коннекционистский метод научения должен быть способным описать создание сети, которая могла бы соответствовать задачам, генерируемым самим мозгом, то есть обрабатывать стимулы внутреннего генеза, без опоры на внешние стимулы;

- нечеткая схема научения: метод научения должен быть достаточно "грубым", чтобы избежать проблем локального минимума, осцилляции и катастрофического забывания, пересказа или потерянных воспоминаний. Некоторые сторонники моделирования психологической системы полагают, что такие проблемы нередко возникают в ходе изучения "естественного" мозга, подверженного заболеваниям. С точки зрения авторов, нет необходимости создавать модели "больных" или неэффективных обучающихся устройств при создании теоретических моделей;

- эффективность научения. Метод должен быть вычислимо эффективным в рамках конечного числа образцов. Необходимо иметь возможность создавать и обучать сеть при условии, что время научения (и создания сети и ее обучения) должно соответствовать полиномиальной функции от числа даваемых примеров;

- обобщение в ходе научения: метод должен позволять осуществлять модификацию сети в ходе научения так, чтобы добиваться наименьшей размерности, то есть результатом научения должна быть наименьшая сеть, способная решать задачу, что позволит избежать проблемы ограниченности ресурсов мозга.

Несмотря на то, что во многих работах посвященных коннекционизму нередко происходит смешение нейронных и коннекционистских сетей, да и значительная часть работ появляется

именно в русле нейронаук, коннекционистский подход применяется и в других областях. В частности методически близкую к коннекционизму позицию занимает известный исследователь вопросов, связанных с обработкой дискурса, В.Кинч (Kintsch, 1988). Его модель понимания основана на вычислительных процедурах, которые приводят к формированию репрезентации текста. Проблема выявления всех когнитивных процессов, которые могут быть вовлечены в конструирование всех правил и структурных элементов репрезентации, сохраняя при этом гибкость достаточную для того, чтобы учитывать все особенности контекста, хорошо известна в психологии речи и обработки текста. Альтернативный подход состоит в привлечении системы продукции "поверхностных" грубых правил, которая генерирует на первом этапе дополнительный контекст, содержащий противоречащие и нерелевантные знания. На втором, интеграционном, этапе система исключает все неподходящие элементы и в результате формируется текстовая база данных, содержащая концепты и пропозициональные элементы, которые извлекаются из данных лингвистического входа несколькими способами. В процессе конструирования используется база знаний субъекта, которая представляется в виде ассоциативной сети, в которой каждый элемент теоретически связан со всеми остальными. Для каждой пары элементов полученной в результате обработки текста базы знаний задана сила связи, которая представляется числом из интервала $[0,1]$. Процесс понимания, как это следует из модели, имеет математическое представление коннекционистского вида, так как и текст на входе системы и в конечном итоге сама система обработки текста имеют вид сети из конечного числа элементов, которые меняют силу своих связей в ходе обработки текста. Несмотря на то, что математическая основа модели Кинча оставляет ряд неясных вопросов, экспериментальная проверка позволяет заключить, что обнаружение противоречий в тексте испытуемыми имеет вид, который предсказывается моделью.

На примере модели Кинча можно видеть пути, по которым идет взаимодействие субсимволического и символического подходов.

Во-первых, следует отметить, что коннекционистские работы будут затрагивать все более высокие уровни когнитивных процессов, что обусловлено эффективностью сетевого представления как такового, а переход от смысловых отношений к различ-

ным формам количественного выражения связей может привести к созданию предпосылок для слияния коннекционистского и символического подходов.

Во вторых, там, где о слиянии пока не может быть речи, все большее распространение получают так называемые гибридные модели, которые используют достоинства моделей традиционной когнитивистской ориентации при изучении психологических явлений внутреннего генеза и коннекционистской, которая имеет несомненное преимущество при изучении процессов распознавания сигналов или образов. Барнден (Barnden, 1994) предлагает разновидность коннекционистских систем, способных осуществлять вычисления, которые моделируют рассуждение по аналогии. По его мнению, это позволит ввести в поле коннекционизма исследования, связанные с продуктивностью.

Основная идея гибридных систем в том, что традиционные модели наиболее эффективно раскрывают особенности процессов сверху - вниз (top-down), которые включают использование систем продукции при решении задач, принятии решений и других сложных процессах, а коннекционистские модели приспособлены для решения процессов идущих снизу - вверх (bottom-up). При обучении, например, иностранному языку, пребывание в естественной языковой среде облегчает настройку коннекционистских сетевых "фильтров", которые облегчают распознавание речи и ее сегментацию, в то же время происходит использование правил уже изученного языка, перенос которых - как успешный, так и неуспешный - происходит в соответствии с традиционными моделями эксплицитного научения.

Один из плодотворных способов "сращения" коннекционистских сетей с модульными системами обработки символической информации содержится в концептуальной схеме устройства, получившего название "черная доска" (blackboard). Это модель, которая допускает асинхронную и автономную активацию баз знаний, продуцирующих множество гипотез, записываемых на "доску". Независимое место в архитектуре этой системы занимает модуль контроля. В одном из вариантов этой системы (Nii, 1986) она включает систему продукции и перцептивный анализатор, которые организованы в сеть. Выходной слой коннекционистской модели анализатора интегрирован в базу знаний генератора новых сетей, что является плодотворным ходом, объединяющим субсимволический и символический уровни.

В своей работе мы также пытаемся соединять преимущества обоих подходов и полагаем, что при введении понятий лингвистической неопределенности или категориальной нечеткости коннекционистская модель является не только эффективным средством реализации процесса обработки нечеткой информации, но и позволяет рассматривать этот процесс с привлечением понятий порядка и беспорядка, которые возникают на субсимволическом уровне в самой коннекционистской системе. Такой взгляд на соотношение уровней помогает глубже понять взаимодействие текстуального сообщения с индивидуальной базой знаний субъекта, а также вскрыть особенности понимания противоречивых и неопределенных составляющих лингвистических стимулов (Kawamoto, 1993).

В завершение вернемся еще раз к проблеме субсимволического и символического уровней когнитивной системы, которые отличаются в первую очередь тем, что первый уровень как бы лишен рефлексивного "осознанного" контроля, а второй не пропускает этот контроль на уровне элементарных "бессмысленных" элементов коннекционистской сети. Мы полагаем, что в пылу критического обострения позиций многие авторы упускают понятие пластичности психологических процессов, способности системы к самоорганизации и функциональной организации, когда функция контроля определяется попеременно стимулами среды и стимулами внутреннего генеза.

Литература

Величковский Б.М. Современная когнитивная психология. М.: МГУ, 1982.

Фодор Дж., Пылишин З. Коннекционизм и когнитивная структура: критический обзор. В кн. Язык и интеллект. М.: Прогресс, 1995.

Abdi H. (1993). *Precis de connexionisme*. In J.F. Le Ny (Ed.), *Intelligence naturelle et intelligence artificielle*, Paris: PUF.

Barnden J.A. (1994). On using analogy to reconcile connections and symbols. *IN: Neural networks for knowledge representation and inference*. D. S. Levine, M. Aparicio IV, (Eds.), Hillsdale: Erlbaum, p. 27-64.

Bechtel, W. (1993). Currents in connectionism.

Chalmers, D.J. (1993). Connectionism and compositionality: Why Fodor and Pylyshyn were wrong.

Fodor J.A. & Pylyshyn Z.W. (1988). Connectionism and cognitive architecture: A critical analysis.

Kawamoto, A.H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account.

Kintsch W. (1988). The role of knowledge in discourse comprehension: A construction - integration model.

Marinov M.S. (1993). On the spuriousness of the symbolic/subsymbolic distinction.

McQueen J.M., Norris D. and Cutler A. (1994). Competition in spoken word recognition: Spotting words in other words

Nii H.P. (1986). Blackboard systems: the blackboard model of problem solving and the evolution of blackboard architectures.

Rosenblatt F. (1961). The perceptron: A probabilistic model for information storage and organization in the brain.

Roy, A., Govil, S. & Miranda, R. (1995). A Neural Network Learning Theory and a Polynomial Time RBF Algorithm. IEEE Transactions on Neural Networks.

Rumelhart D.E., and McClelland J.L. (1986).

Rumelhart D.E., Hinton G.E. & Williams R.J. (1986). Learning internal representations by error propagation. In Rumelhart D.E., and McClelland J.L. (Eds.). *Parallel distributed processing*. Cambridge: MIT Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1-74.

Widrow B. & Hoff M.E. (1960). Adaptive switching circuits. 1960 IRE WESCON Convention Records, 96-104.