Development of Cue Integration with Reward-mediated Learning

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften

vorgelegt beim Fachbereich Informatik und Mathematik
der Goethe-Universität
Frankfurt am Main

von

Thomas Henning Weißwange
aus Mainz

Frankfurt am Main, 2011

(D30)

vom Fachbereich Informatik und Mathematik der Goethe-Universität Frankfurt am Main als Dissertation angenommen.

Dekan: Prof. Dr. Tobias Weth

1. Gutachter: Prof. Dr. Jochen Triesch

2. Gutachter: Prof. Dr. Visvanathan Ramesh

Datum der Disputation: 04.06.2012

Zusammenfassung

Eine wichtige Voraussetzung für das Überleben eines jeden Tieres ist seine Fähigkeit auf seine Umgebung zu reagieren. Um dazu in der Lage zu sein, muss es versuchen möglichst viele Informationen über den aktuellen Zustand der Umwelt zu sammeln. Hierzu stehen ihm eine Reihe, von Tierart zu Tierart unterschiedliche, Sinnesorgane zur Verfügung. Ein Problem ergibt sich jedoch aus der Tatsache, dass alle diese Sensoren für die meisten Variablen nicht die direkten Werte messen, sondern nur eine niedrigdimensionale Transformation selbiger liefern können. Eine solche transformierte Messung ist nun aber mehrdeutig gegenüber des ursprünglichen Zustands, das heißt verschiedenen Werte derselben können zur gleichen Messung führen.

Wenn das Auge zum Beispiel einen länglichen gelben Fleck zwischen einer Menge grün signalisiert, könnte das darauf hinweisen, dass sich ein Tiger im Gebüsch versteckt. Es ist allerdings genauso gut möglich dass es sich bei dem Fleck nur um eine Banane, oder einen Vogel handelt. Um nicht ständig vor Bananen davon zulaufen, und damit wertvolle Energie zu verschwenden, sollte ein Tier daher versuchen die Mehrdeutigkeiten einzugrenzen. Dazu gibt es mehrere Möglichkeiten, zum Beispiel könnte es ein wenig länger warten und sehen ob sich der Fleck vielleicht bewegt, was die Tiger Erklärung deutlich wahrscheinlicher machen würde. Alternativ könnte der Kontext der Messung miteinbezogen werden, etwa auf welcher Höhe sich der Fleck befindet oder ob Bananen in der aktuellen Jahreszeit überhaupt schon reif sein können. Ein weiterer Ansatz, und derjenige mit dem sich der größte Teil dieser Arbeit befassen wird, ist die Verwendung eines zweiten unabhängigen Sinnesorgans. Hört man zum Beispiel zur gleichen Zeit knackende Äste, erhöht das wieder die Wahrscheinlichkeit eines Tigers, obwohl das Geräusch für sich allein genommen auch von vielen anderen Dingen hätte ausgelöst werden können. Diese letztgenannte Variante bezeichnet man als “Multimodale” oder “Multisensorische Integration”.

Die Integration verschiedener Sinnesorgane beschäftigt die moderne Wissenschaft, speziell die Psychophysik, bereits seit knapp 100 Jahren (e.g. [Todd 1912]). Die oben beschriebene Intuition der Verbesserung mit der Hinzunahme einer weiteren Informationsquelle bestätigten sich auch in quantitativen Experimenten. Der Effekt zeigt sich in verschiedenen Aspekten: **I.** Einer Verringerung der Reizschwelle, zum Beispiel können Lichtblitze mit geringem Kontrast nur detektiert werden wenn sie von einem kurzen Geräusch begleitet werden. **II.** Einer Verbesserung der Genauigkeit, wodurch man zum Beispiel deutlich geringere Größenunterschiede zweier Objekte erkennen kann wenn man sie nicht nur sehen sondern auch berühren darf.

Auf der Suche nach einer Erklärung für diese und andere psychophysikalische Experimente, hat es sich als sehr hilfreich erwiesen, den Prozess der Wahrnehmung der Umwelt als probabilistischen Prozess zu betrachten. Jede sensorische Messung verändert die Wahrscheinlichkeits-Verteilung über deren mögliche Ursachen in der Welt. Die theoretische Basis für Berechnungen mit solchen bedingten Wahrscheinlichkeiten (die Wahrscheinlichkeit einer Ursache bedingt auf die spezifische Sensor-Antwort) ist Bayes’ Theorem. Grob fromuliert angewandt auf das obige Beispiel ergibt sich daraus, dass die Wahrscheinlichkeit eines Tigers, gegeben dass wir einen gelben Fleck sehen (“A-posteriori Wahrscheinlichkeit”), proportional zum Produkt der Wahrscheinlichkeiten ist, dass ein Tiger einen solchen gelben Fleck verursacht (“Likelihood”), und dass in der aktuellen Umgebung Tiger existieren (“A-priori Wahrscheinlichkeit”). Die Verwendung dieser Formel garantiert, dass alle vorhandenen Information in optimaler Art und Weise benutzt werden.

Neuere experimentelle Studien konnten zeigen, dass der Mensch tatsächlich in der Lage ist, Leistungen zu erbringen, die quantitativ ähnlich den Vorhersagen eines solchen Bayes’schen Modells sind. Die meisten dieser Arbeiten befassen sich mit Multisensorischer Integration, so zum Beispiel mit der bereits erwähnten visuo-haptischen Größen-Unterscheidung, aber auch mit der Kombination von verschiedenen Informationsquellen innerhalb eines einzelnen Sinnesorgans wie der Tiefenschätzung mittels Stereo- und Bewegungs-Informationen. In solchen Experimenten wird als erstes die Genauigkeit der Antworten für jeden Sinn allein gemessen, und aus diesen Werten kann dann mit Hilfe der Bayes Formel eine Vorhersage über die optimale Kombination im multisensorischen Fall berechnet werden. Zum Vergleich werden dann multisensorische Stimuli gezeigt, bei denen allerdings den beiden Sinnen jeweils unterschiedliche Werte

gezeigt werden, ohne dass dies den Testpersonen bewusst ist. Anhand der Tendenz in Richtung des einen oder anderen Wertes kann nun die Gewichtung der beiden Informationsquellen errechnet und mit den Vorhersagen verglichen werden. Die durchschnittlichen Gewichte aller Testpersonen stimmen in vielen dieser Studien mit denen des Bayes'schen Modells überein.

Bei der Wiederholung ähnlicher Experimente mit Säuglingen und Kindern verschiedener Alters-Stufen wurde kürzlich jedoch festgestellt, dass diese Fähigkeit erst im Laufe der Entwicklung entsteht. Abhängig von der getesteten Aufgabe zeigten sich deutliche Abweichungen vom Bayes'schen Modell bis zum Alter von 10 Jahren. Die Ergebnisse waren eindeutig weder das Resultat von Schwächen in den einzelnen Sinnesorganen, noch bedingt durch einen eventuell fehlenden anatomischen Weg die beiden Informationen zusammen zu bringen.

Dies ist auch deshalb interessant, weil existierende Theorien zu Multisensorischer Integration oft auf der Annahme basieren, dass die im Gehirn verwendeten Strukturen explizit probabilistische Eigenschaften haben. Diese theoretischen Modelle lassen sich grob, der von David Marr [Marr 1982] vorgeschlagenen Nomenklatur folgend, auf zwei Beschreibungsebenen aufteilen[1]:

Theorien der algorithmischen Ebene beschäftigen sich vorrangig damit Approximationen für Bays'sche Berechnungen zu finden, die es dem Gehirn ermöglichen würden, Ergebnisse innerhalb einer akzeptablen Zeit zu bekommen. Die vollständige numerische Berechnung der Wahrscheinlichkeitsverteilungen stellt sich für reale Aufgaben aufgrund der hohen Dimensionalität als unlösbar dar. Eine einfache und effiziente Approximations-Möglichkeit, speziell für Multisensorische Integration, ist das Bayes'sche gewichtete Mittel (wie es auch in den oben erwähnten Psychophysik Experimenten zum Vergleich verwendet wird). Dabei wird für jeden Sinn separat ein Schätzwert berechnet und diese Werte dann gewichtet gemittelt, wobei die Gewichte proportional zur Genauigkeit des jeweiligen Sinnes sind. Ein Nachteil dieser Methode ist, dass sie nur dann korrekt ist, wenn die beteiligten Informationsquellen strikte Eigenschaften erfüllen. Diese Eigenschaften kann man zwar im Labor kontrollieren, im realen Alltag zeigt sich allerdings, dass diese nicht immer erfüllt sind.

Die niedrigste Beschreibungsebene befasst sich mit der möglichen Implementierung im Gehirn. Ein bekanntes Modell dafür, genannt Probabilistic Population Code (PPC), besteht aus neuronalen Recheneinheiten mit probabilistischer Feuerrate basierend auf der Likelihood eines Stimulus'. Ist eine Gruppe solcher Einheiten gleichmäßig über den Raum der Möglichen Werte verteilt (d.h. durch die entsprechenden Likelihood-Funktionen bedeckt), so repräsentiert die Höhe des Ausschlags der Aktivität der gesamten Population die Varianz der A-posteriori Verteilung. Möchte man nun zwei solcher Populationen kombinieren, die zum Beispiel aus zwei verschiedenen Sensoren gespeist werden, so reicht eine simple Addition der jeweiligen Einheiten die den gleichen Werte-Bereich repräsentieren, um die zwei Sensoren optimal zu integrieren. Dieses Modell bezieht sich direkt auf die zuvor erwähnte Approximation durch ein Bayes'sches gewichtetes Mittel, und gilt somit nur für eine begrenzte Zahl an Verteilungen. Noch problematischer ist allerdings, dass dieses, wie auch andere Modelle, keinerlei Entwicklung im Sinne der psychophysikalischen Ergebnisse (s.o.) zulässt. Die Fähigkeit zur fast-optimalen Integration basiert auf intrinsischen Eigenschaften der Neurone und damit auf etwas, dass schon bei Geburt vorhanden ist.

Das Hauptanliegen dieser Doktorarbeit ist es, ein alternatives Modell zu entwickeln, welches sowohl den Optimalitäts- als auch den Entwicklungsaspekt von Multisensorischer Integration abdecken kann. Dabei soll zuerst ein algorithmisches Prinzip auf sein Potential getestet werden, und dieses danach in ein detailliierteres Modell einer möglichen Implementierung übertragen werden. Das Lernparadigma, auf dem beide Modelle aufbauen, ist Reinforcement Learning (RL). Ursprünglich als theoretische Methode zur Lösung von Markov Decision Problems (MDPs) entwickelt, konnten spätere experimentelle Studien überzeugende Hinweise darauf liefern, dass RL auch bei Lernvorgängen von Tieren und Menschen eine wichtige Rolle spielt. Mechanistisch steht RL zwischen klassischem überwachten Lernen, bei dem ein "Lehrer" die jeweils richtigen Antworten liefert, und unüberwachtem Lernen, welches rein aus statis-

[1] Die dritte und höchste Ebene der theoretischen Beschreibung ("*computational theory*") ist bereits mit der Definition von Wahrnehmung als probabilistischem Prozess, den es zu optimieren gilt, abgedeckt.

tischen Mustern in den Daten lernt. In RL definiert man meist einen Zustand der Umgebung und darauf basierend eine Aktion des “Lerners” (auch “Agent”). Basierend auf der Qualität dieser Aktion wird ein positives oder negatives Lern-Signal (“Reward”[2]) gegeben. Auch diese ursprüngliche technische Formulierung war schon inspiriert vom Verhalten biologischer Organism, welche häufig auch nur aus den Resultaten ihrer Handlungen lernen können ohne jedoch die korrekte Reaktion vorgegeben zu bekommen. Eine der wichtigsten Algorithmen innerhalb von RL ist das so genannte “Temporal Difference” (TD) Lernen: Basierend auf vorangegangenen Erfahrungen macht der Agent eine Vorhersage über den erwarteten Reward für jeden der möglichen Aktionen im aktuellen Zustand der Umgebung. Diese Vorhersage kann zusätzlich auch mögliche spätere Rewards miteinschließen, welche sich aus dem einer Aktion folgenden (veränderten) Zustand ergeben könnten. Der Agent wählt basierend auf diesen Vorhersagen eine Aktion zur Ausführung aus (zum Beispiel diejenige mit der höchsten Vorhersage) und erhält daraufhin einen Reward. Diesen vergleicht er nun mit seiner Vorhersage und verwendet den Fehler um seine zukünftigen Vorhersagen zu verbessern. Ein solcher Algorithmus konvergiert, unter einigen Voraussetzung, gegen die optimale Lösung.

Aufgrund dieser biologischen wie theoretischen Ergebnisse denken wir, dass das RL-Konzept ein guter Kandidat für eine Erklärung des Lernens von Multisensorischer Integration ist.

Das RL nicht nur auf der konzeptuellen sondern auch auf der algorithmischen Ebene biologischem Verhalten ähnelt, konnte wenig später mit elektrophysiologischen Experimenten gezeigt werden. Ableitungen von dopaminergen Neuronen in der *Area tegmentalis ventralis* (ATV) zeigten Korrelationen zwischen dem Muster der Aktionspotenziale und dem theoretischen Vorhersage-Fehler. Es war bereits bekannt, dass diese Neuronen als Antwort auf ein Belohnungs-Signal ihre Feuerrate erhöhen. Die neueren Experimente zeigten aber, dass dieses Verhalten verschwindet, wenn ein Reward vollkommen vorhersehbar ist. Stattdessen fand man eine solche Reaktion aber nun als Antwort auf einen Stimulus (z.B. einen Ton) der diesen Reward ankündigte. In anderen Arealen wurden später Repräsentationen von weiteren wichtigen TD-Lern-Variablen gefunden.

In den letzten Jahren befassten sich auch erste theoretische und experimentelle Arbeiten mit möglichen Implementierungen von RL im neuronalen Substrat des Gehirns. Die Basis all dieser Arbeiten ist der Einfluss von Dopamin, dem neuronalen Substrat eines TD Vorhersage Fehlers, auf synaptische Plastizitäten. Mehrere Publikationen konnten zeigen, dass Dopamin die Ausprägungen von “Spike-Timing-Dependent Plasticity” (STDP) beeinflusst, einem der Haupt-Akteure aller Lernvorgänge im Gehirn. Theoretische Studien konnten eine dazu passende mathematische Regel formulieren bei der das Reward-Signal auf den STDP-Kernel multipliziert wird. In fast allen bisherigen Modellen und Simulationen wurden die Auswirkungen dieses so genannten R-STDP (“Reward-modulated STDP”) in Isolation untersucht. Wir wissen aber, dass im Gehirn viele verschieden Plastizitäts-Mechanismen zusammen wirken, und frühere Untersuchungen unserer Gruppe konnten zeigen, dass komplexe Verhaltens-Muster in Simulationen Neuronaler Netze nur entstehen konnten wenn man mehrere Lernregeln gleichzeitig verwendet. Solche anderen Plastizitäten wirken zum Beispiel an hemmenden Synapsen (STDP wird meist nur an Erregenden Verbindungen verwendet), oder regulieren homeostatisch die Induktionsschwelle für Aktionspotentiale. Im zweiten Teil dieser Arbeit werden wir uns daher damit auseinandersetzen, ob R-STDP im Zusammenspiel mit solchen anderen Plastizitäten in der Lage ist, ein Simuliertes Netzwerk so anzupassen, dass es zu Multisensorischer Integration in der Lage ist. Gleichzeitig können wir die Gegenseitige Einflussnahme der Lernregeln auf einander analysieren.

[2] Im Deutschen häufig mit Belohnung übersetzt. Da ein Reward jedoch sowohl positiv als auch negativ sein kann, verwenden wir hier das englische Wort.

Algorithmisches Model zur Entwicklung von Multisensorischer Integration.

Wir untersuchen, ob ein mit Reinforcement Learning trainiertes Computer-Modell lernt, die Aufgabe zu lösen, aus zwei in ihrer Aussage mehrdeutigen Informations-Quellen den ursprünglichen Wert einer Variable zu berechnen. Um die Äquivalenz dieses Problems mit Multisensorischer Integration zu verdeutlichen verwenden wir im folgenden beispielhaft eine konkretere Formulierung. Ein Agent soll mit seiner Aktion (z.B. einer Greifbewegung) die reale Position (entlang einer räumlichen Dimension) eines Objekts in der Welt schätzen. Dafür bekommt er je eine verrauschte (mittels Gauss'schem Rauschen) visuelle und auditorische Messung. Basierend auf der Güte dieser Aktion wird ein Reward berechnet (z.B. der Erfolg der Greifbewegung). Ein jeder Durchgang besteht also aus einem einzelnen Zeitschritt mit neuer Objekt-Position, audio-visueller Messung, Aktion und Reward/Lernen. Wir verwenden TD-Lernen und setzen ein künstliches neuronales Netz (KNN) ein um die Funktion zu approximieren, welche einen Zustand (hier ein Paar von Messungen) und eine Aktion auf eine Reward-Vorhersage abbildet. Wir verwenden standardmäßige Gradienten-basierte Lernregeln, um das KNN nach jedem Versuch anzupassen, mit der Besonderheit, dass das Lern-Signal nicht aus der korrekten Antwort, sondern aus dem Fehler in der Reward-Vorhersage besteht.

Die Ergebnisse eines solchen Trainings ähneln denen eines Bayes'schen Modells, welches die audio-visuellen Messungen in optimaler Weise kombiniert. Unter anderem finden wir eine Verbesserung der Genauigkeit gegenüber einem optimalen Modell welches nur eine einzige Informations-Quelle verwenden kann. Um die Aufgabe noch zusätzlich zu erschweren, ändern wir die Durchgänge in dem wir zufällig ein oder zwei Objekte positionieren. Das bedeutet für den Agenten dass er aus den selben zwei Messungen zusätzlich die Information extrahieren muss, ob er sie Integrieren muss, oder ob er besser eine der Messungen ignorieren sollte. Bei einer solchen Aufgabe gibt es ebenfalls psychophysikalische Belege für ein Verhalten der Testpersonen, welches den Bayes'schen Vorhersagen sehr ähnlich ist. Auch in diesem Fall konnte unser RL-Agent die experimentellen Ergebnisse reproduzieren. In mehreren weiteren Variationen dieser Aufgabe demonstrieren wir außerdem zum Beispiel die Robustheit des Ansatzes gegenüber Änderungen in der Stärke des Rauschens in den beiden Messungen. Auch die Verwendung von A-Priori ungleich-verteilten Positionen, die Verwendung von nicht-Gauss'schen Rausch-Funktionen oder eines systematischen Fehlers in selbigen kann von dem Netzwerk gelernt werden.

In einer Kollaboration mit dem Honda Research Institute Europe (HRI-EU) konnte das Modell in der Folge auch auf realistischeren Daten und Aufgaben getestet werden. Die beiden dafür verwendeten Datensätze stammen aus auditorischen bzw. visuellen Aufnahmen von Stimuli mit verschiedenen Distanzen von einem Roboter-Kopf. Aus diesen Aufnahmen wurden eine Anzahl unterschiedlicher Informations-Quellen berechnet. Für das auditorische Set sind das zum Beispiel unter anderem der Zeitunterschied der Ankunftszeit eines Tons zwischen den beiden Ohren (ITD: "Interaural Time Difference"), oder die Lautstärke des Tons, für das visuelle Set zum Beispiel die Verschiebung der Bilder in den beiden Kameras gegeneinander. Für jede dieser Quellen kann man die Distanz eines Objekts schätzen. Durch die Integration aller verfügbaren Quellen sollte man damit, nach Bayes, in der Lage sein die Genauigkeit dieser Schätzungen zu verbessern. Da das Rauschen in den Schätzungen der Quellen aber nicht mehr zwingend eine mathematisch simple Verteilung (also nicht mehr z.B Gauss'sch oder unkorreliert ist), wie es in der Simulation oder bei psychophysikalischen Experimenten der Fall ist, wissen wir, dass eine einfach Approximation mittels gewichtetem Mittelwert nicht mehr garantiert die optimale Lösung bringt. Und tatsächlich ist eine solche integrierte Schätzung im visuellen Fall oft sogar schlechter als eine der einzelnen Quellen. Lassen wir stattdessen unseren RL-Agenten die beiden Tiefenschätzungs-Aufgaben lernen, so finden wir eine deutliche Verbesserung der Ergebnisse durch die Integration mehrerer Quellen. Der Algorithmus ist zum Beispiel in der Lage sich an Tiefenabhängige Qualitätsunterschiede innerhalb einer einzelnen Quelle anzupassen, oder Korrelationen und systematische Fehler auszugleichen.

Diese Ergebnisse sind damit potentiell also auch für Anwendungen, zum Beispiel in der Robotik, interessant, da wir nicht nur zeigen konnten, dass die Methode besser und robuster sein kann als eine einfach Approximation von Bayes'scher Integration, sondern auch weil eine einzelne Integration nach Abschluss des Lernens effizient berechnet werden kann (Ein einziger Durchlauf durch das KNN).

Insgesamt kommen wir zu dem Ergebnis, dass Reward-basiertes Lernen in der Lage ist ohne zusätzliche Annahmen oder Informationen über die Umgebung eine Leistung zu erzielen die der eines optimalen Bays'schen Modells sehr nahe kommt. Zusätzlich dazu kann ein solcher Agent auch lernen auf zusätzliche versteckte Variablen, wie die Anzahl an Objekten, in fast-optimaler Art und Weise zu reagieren. In Kombination mit den experimentellen Studien über RL im Gehirn denken wir daher, dass Reinforcement Learning eine plausible Möglichkeit darstellen kann, um den Entwicklungs-Prozess eines Kindes hin zu fast-optimaler Multisensorischer Integration zu erklären. Dies schließt allerdings nicht aus, dass sich zusätzlich auch noch weitere Mechanismen an diesem Prozess beteiligen könnten, oder dass Erwachsene später auch andere Methoden verwenden könnten. Nichtsdestotrotz ist dies die erste theoretische Arbeit, die sich mit dem Entwicklungs-Prozess Multisensorischer Integration beschäftigt, und sie soll vor allem auch konstruktiv auf Schwächen der bisher dominierenden Thesen hinweisen.

Modell zur Implementierung von Multisensorischer Integration.

Nachdem wir zeigen konnten, dass ein RL-basierender Algorithmus generell in der Lage ist, Multisensorischer Integration zu erlernen die ähnliche Ergebnisse liefert wie ein optimales Modell, wollen wir mögliche neuronale Implementierungen eines solchen Algorithmus' untersuchen. Um die Komplexität zu begrenzen verwenden wir ein simples deterministisches Neuronen-Modell, welches zwar, wie sein biologisches Äquivalent, binär über Aktionspotentiale kommuniziert, allerdings keinerlei Gedächtnis besitzt (es summiert nur alle gewichteten zur gleichen Zeit einkommende Signale und vergleicht diesen Wert mit einer Reizschwelle). Eine Population solcher spärlich untereinander verbundener Neuronen (ein so genanntes "Reservoir") erhält externe Stimulation durch auditorische und visuelle Messungen[3] und wird dann für mehrere diskrete Zeitschritte simuliert. Eine separate Gruppe an Neuronen (Ausgabe-Neurone) liest die Aktivität des Reservoirs zu einem bestimmten Zeitpunkt aus und bestimmt durch ihre Aktivität die auszuführende Aktion. Die Aufgabe ist die gleiche wie im vorherigen Kapitel, mit dem einzigen Unterschied, dass jetzt auch eine zeitlich Komponente vorhanden ist (z.B. wie lange gewartet wird bis eine Aktion ausgeführt werden muss). Wir verwenden mehrere verschiedene Plastizitäts-Mechanismen an verschiedenen Verbindungen und Neuronen um das Netzwerk an diese Aufgabe zu adaptieren. Allen gemein ist jedoch, dass sie auf experimentellen elektrophysiologischen Daten beruhen. In früheren Arbeiten unserer Gruppe konnte bereits gezeigt werden, dass erst eine Kombination mehrerer Plastizitäten in der Lage ist komplexere Aufgaben mit einem solchen simplen Netzwerk zu lösen. Basierend darauf verwenden wir auch hier "Spike-timing-dependent Plasticity" (STDP) zusammen mit "Synaptic Scaling" (oder Multiplikative Normalisierung) an erregenden Synapsen und einen homeostatischen Mechanismus (IP) zur Regulierung der Reizschwelle der Neuronen. Zusätzlich fügen wir in dieser Arbeit eine Plastizität an hemmenden Synapsen (iSTDP) ebenfalls kombiniert mit Synaptic Scaling ein. Die erwähnten Mechanismen gelten innerhalb des Reservoirs, für die Verbindungen zu den Ausgabe-Neuronen verwenden wir Reward-moduliertes STDP (R-STDP), wobei der TD-Vorhersage-Fehler als modulierendes Signal eingesetzt wird.

Wir können zeigen, dass dieses Netzwerk nach dem Training eine größere Genauigkeit besitzt, als mit nur einer Informations-Quelle möglich wäre. Da das gleiche Netzwerk vor dem Training (zufällig initialisiert) nicht in der Lage ist die Aufgabe zu bewältigen, können wir sagen, dass das Netzwerk Multisensorische Integration erlernt hat. Bedauerlicherweise zeigt ein Vergleich mit dem optimalen Modell, dass die Leistung des Netwerks doch deutlich darunter liegt. Der Hauptgrund dafür liegt allerdings nicht

[3] Wir verwenden weiterhin die bildliche Nomenklatur aus dem vorherigen Kapitel.

im Reward-modulierten Ausgabe-Lernen - ein Vergleich mit einem offline-trainierten überwachten Lernverfahren für die Ausgabe zeigt keine oder nur sehr geringe Unterschiede. Stattdessen scheint es dem Reservoir an Gedächtnis zu fehlen, denn je länger die Pause zwischen einem Stimulus und der Aktion ist desto schlechter schneidet das Netzwerk ab. Da die Neurone selbst kein Gedächtnis besitzen müsste ein solches aus der Dynamik des Netzwerks selbst entstehen. Vorgänger-Studien haben gezeigt, dass dies durchaus möglich ist und durch den Einsatz von STDP und IP im Reservoir verstärkt werden kann. Allerdings unterscheiden sich die Eingabe-Statistiken in unserem Fall doch deutlich, da zum Beispiel zusätzliches Rauschen und auch insgesamt mehr Stimuli verwendet werden.

Trotz dieser Limitierungen erkennt man deutlich die Einflüsse der Reservoir-Plastizität, denn ein statisches Reservoir zeigt für fast alle Initialisierungen keine oder nur schlechte Integration. Entfernt man IP aus einem plastischen Reservoir, so führt dies zumeist zu uniformen Netzwerk-Aktivitäten, das heißt häufig sind alle oder keine Neuronen aktiv. In diesem Fall verringert sich damit auch die Gedächtnisspanne nochmals, und das Netzwerk enthällt generell nur noch wenig Information. Ähnlich stark wirkt sich das Entfernen von iSTDP aus. Meist entstehen dadurch Oszillation, mit wechselnden Phasen hoher und niedriger Netzwerk-Aktivität, wiederum mit sehr negativen Auswirkungen auf die Leistung. Interessanterweise finden wir keine Veränderung in der Netzwerk-Funktion, wenn wir STDP ausschalten. Dies steht im Gegensatz zu einigen der früheren Arbeiten, und verdeutlicht, dass die Interaktion mehrerer Plastizitäts-Mechanismen auch sehr von der Struktur der Eingaben abhängt.

Unsere Ergebnisse haben gezeigt, dass es möglich ist mit einem biologisch plausiblen Reward-modulierten Plastizitäts-Mechanismus zu lernen mehrere Informations-Quellen zu integrieren. Um zu testen ob dies in einer Qualität erreichbar ist, die der Vorgabe eines optimalen Bayes'schen Modells nahe kommt, wird jedoch ein leistungsfähigeres Reservoir benötigt. Die Ergebnisse mit dem hier verwendeten simplen Neuronen-Modell eignen sich trotzdem dafür Aussagen über das Zusammenspiel der unterschiedlichen Plastizitäten zu treffen. Neu dabei ist vor allem der Effekt der Lernregel für hemmende Synapsen, welcher noch kaum untersucht ist und in unserem Falle gemeinsam mit IP sehr gut die Netzwerk-Dynamiken zu kontrollieren scheint. Besonders interessant dabei ist die Verbindung zu den theoretischen Hintergründen von IP - die Formulierung der Lernregel zielte explizit darauf ab, die "Sparseness" der Feuerrate einzelner Neurone zu optimieren, so dass bei geringer mittlerer Aktivität Informationen bestmöglich weitergegeben werden. "Sparseness" konnte aber experimentell nicht nur für das Feuern eines Neurons innerhalb eines längeren Zeitfensters, sondern auch für die Aktivität einer größeren Population innerhalb eines Zeitschrittes beobachtet werden. In unseren Simulationen scheint die Kontrolle dieser letzteren Form über die Veränderung der hemmenden Synapsen mittels iSTDP stattzufinden. Diese Ergebnisse könnten sich auch auf Netzwerke mit komplexeren Neuronen-Modellen übertragen lassen, und damit Hinweise darauf bringen wie das Gehirn durch solche lokalen Lernregeln globale Aufgaben lösen kann.

Insgesamt können wir mit dieser Arbeit zeigen das Reinforcement Learning eine plausible Methode sein kann, um eine Entwicklung von fast-optimaler Multisensorischer Integration zu erklären. Im Gegensatz zu existierenden Theorien und Modellen benötigen wir dazu keinerlei Annahmen über spezifische probabilistische Eigenschaften innerhalb der neuronalen Strukturen.

Contents

1
Motivation and Thesis Overview

A central requirement for the survival of every living being is to know about the states of its environment. In particular, it wants to infer the state of certain variables that are important for its survival. To do so, it has to rely on the signals (also called "cues") that it gets from its sensory receptors. But usually those receptors do not sense the value of the relevant variables directly. To use an example, hearing a noise from behind could mean that there is a tiger in the bushes, but it could also just be the wind moving leafs or any of a number of other less critical events. The activation of a receptor can be seen as influencing the probability for a variable taking a certain value, hearing a noise increases the probability for tigers, wind and so on. Intuitively getting information from an additional, independent, receptor will improve your estimates. In our example, also seeing patches of yellow will make the tiger explanation more likely, although on its own this could as well be caused by a banana. This principle of using information from multiple sensors to improve behaviour can be found everywhere, from the human brain [Thomas 1941] down to single cell level [Adler & Tso 1974, Khan *et al.* 1995]. Interestingly, it is not yet clear how much of the high level behaviour results directly from cellular features and how much is the result of a developmental/learning process.

The theoretical framework for computing with conditioned probabilities, like that of a tiger being present given a certain cue, is based on Bayes' Theorem. Using it for our example, it would state that the probability that a tiger is present given a noise, is proportional to the product of the probability that a tiger elicits such noise and the general probability of tigers in the environment. This formulation guarantees to make optimal use of all available information. In that line of thinking a second source of information about the probability of a tiger should help improving decisions.

Integration of multiple senses has been a topic of modern psychophysics research for almost a century (e.g. [Child & Wendt 1938, Thomas 1941, Todd 1912]), and it was found that it can improve precision and reaction time over that of unisensory stimuli. Only relatively recently though, scientists started to compare human and animal performance with predictions of optimal models based on Bayes' theory [Geisler 2011], and did often find good agreements between the two [Alais & Burr 2004, Battaglia *et al.* 2003, Ernst & Banks 2002, Jacobs 1999, Knill & Saunders 2003].

These findings ask for a theoretical approach on how the brain is able to produce such near-optimal behaviour. Some models of probabilistic computation in the brain have been proposed (for an overview see [Doya *et al.* 2007, Knill & Pouget 2004]), that use for example a population of probabilistic spiking neurons to encode distributions [Ma *et al.* 2006, Ma *et al.* 2008] or fire with a rate that is a function of the probability of an event [Deneve 2008a, Gold & Shadlen 2001]. These models imply that representing and computing with probability distributions should be an intrinsic property of the neural code. If those assumptions are true, it could follow that the brain should be able to act "Bayesian" from birth on or show no integration at all until all anatomical connections are established and than suddenly perform near-optimal. Experiments of developmental psychologists recently showed that this is not the case. Instead the ability to integrate cues with close to optimal performance is not present at birth but develops over time (taking from a few months up to many years) [Gori *et al.* 2008, Nardini *et al.* 2008, Nardini *et al.* 2010, Neil *et al.* 2006].

This thesis targets the question of how such close to optimal behaviour can develop with experience, both on a more conceptual as well as on an implementational level. I use as few preliminary assumptions as possible and show that interaction with the environment is an important factor for learning. In contrast to previous modelling studies, I will not use any methods or assumptions from Bayesian theory but instead show that similar results can be produced by a model-free reward based learning scheme.

Thesis overview

This thesis will first introduce in more detail the Bayesian theory and its use in integrating multiple information sources. I will briefly talk about models and their relation to the dynamics of an environment, and how to combine multiple alternative models.

Following that I will discuss the experimental findings on multisensory integration in humans and animals. I start with psychophysical results on various forms of tasks and setups, that show that the brain uses and combines information from multiple cues. Specifically, the discussion will focus on the finding that humans integrate this information in a way that is close to the theoretical optimal performance. Special emphasis will be put on results about the developmental aspects of cue integration, highlighting experiments that could show that children do not perform similar to the Bayesian predictions. This section also includes a short summary of experiments on how subjects handle multiple alternative environmental dynamics. I will also talk about neurobiological findings of cells receiving input from multiple receptors both in dedicated brain areas but also primary sensory areas.

I will proceed with an overview of existing theories and computational models of multisensory integration. This will be followed by a discussion on reinforcement learning (RL). First I will talk about the original theory including the two different main approaches model-free and model-based reinforcement learning. The important variables will be introduced as well as different algorithmic implementations. Secondly, a short review on the mapping of those theories onto brain and behaviour will be given. I mention the most influential papers that showed correlations between the activity in certain brain regions with RL variables, most prominently between dopaminergic neurons and temporal difference errors. I will try to motivate, why I think that this theory can help to explain the development of near-optimal cue integration in humans.

The next main chapter will introduce our model that learns to solve the task of audio-visual orienting. Many of the results in this section have been published in [Weisswange *et al.* 2009b, Weisswange *et al.* 2011]. The model agent starts without any knowledge of the environment and acts based on predictions of rewards, which will be adapted according to the reward signaling the quality of the performed action. I will show that after training this model performs similarly to the prediction of a Bayesian observer. The model can also deal with more complex environments in which it has to deal with multiple possible underlying generating models (perform causal inference). In these experiments I use different formulations of Bayesian observers for comparison with our model, and find that it is most similar to the fully optimal observer doing model averaging. Additional experiments using various alterations to the environment show the ability of the model to react to changes in the input statistics without explicitly representing probability distributions. I will close the chapter with a discussion on the benefits and shortcomings of the model.

The thesis continues whith a report on an application of the learning algorithm introduced before to two real world cue integration tasks on a robotic head. For these tasks our system outperforms a commonly used approximation to Bayesian inference, reliability weighted averaging. The approximation is handy because of its computational simplicity, because it relies on certain assumptions that are usually controlled for in a laboratory setting, but these are often not true for real world data. This chapter is based on the paper [Karaoguz *et al.* 2011].

Our second modeling approach tries to address the neuronal substrates of the learning process for cue integration. I again use a reward based training scheme, but this time implemented as a modulation of synaptic plasticity mechanisms in a recurrent network of binary threshold neurons. I start the chapter with an additional introduction section to discuss recurrent networks and especially the various forms of neuronal plasticity that I will use in the model. The performance on a task similar to that of chapter 3 will be presented together with an analysis of the influence of different plasticity mechanisms on it. Again benefits and shortcomings and the general potential of the method will be discussed.

I will close the thesis with a general conclusion and some ideas about possible future work.

2
Introduction

Human perception has to deal with great ambiguity within its inputs. The original state of the world is transformed by receptors that often reduce or transform the dimensionality of the stimuli. The eye for example maps 3D objects to the activity on the 2D grid of the retina. The same retinal image could be caused by a large number of 3D structures (Fig. 2.1). It is now widely accepted to look at perception as a probabilistic process [Bülthoff & Yuille 1991, Doya *et al.* 2007, Kersten & Yuille 2003, Knill & Pouget 2004, Perfors *et al.* 2011] (see also a special issue of Trends in Cognitive Sciences [Chater *et al.* 2006]). The general idea that the brain tries to infer the underlying states from ambiguous perception can be traced back to Al Hazen around the year 1000 [Smith 2001] and later to von Helmholtz [von Helmholtz 1867] in the 19th century. A probability can be assigned to all the structures from the above example, depending on how reliably they will cause the exact image. Additional prior knowledge on the frequency of these structures appearing in the current context can be used to refine the distribution over possible causes [Mamassian & Landy 1998]. Another way to improve the estimates is to use measurements of the same structure from additional sensors.

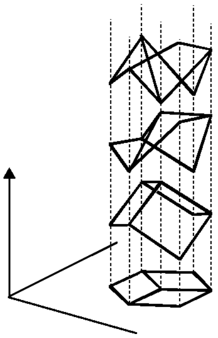

Figure 2.1. A specific object shape in the 2D plane could be caused by multiple different 3D objects (Adapted from [Sinha & Adelson 1993], ©1993 IEEE).

When one talks about theoretical models of cognitive processes it is useful to adopt the notion of the three levels of description introduced by David Marr [Marr 1982]. The above description of perception can be considered to address the highest level, that of *computational theory.* It tries to explain the goal of perception, which as I said before seems to be inference of the true state of the world from ambiguous inputs. Bayesian theory, which will we be the topic of the next section, is the mathematical formulation for the upper limit of performance for those goals. Therefore it is interesting to compare these limits with the ability of the brain to solve inference tasks. As we will see later, the Bayesian formulation showed to be a good match for human perceptual performance, at least in a subset of tasks.

The middle level of description, referred to as the *algorithmic level*, addresses the mechanisms that are used by the brain to solve the computational task. One possibility would be a direct use of the Bayesian equations, or at least the implementation of approximations of it, e.g. sampling, due to the enormous complexity of inference processes in natural tasks. Theories on this level are often tested or developed from psychophysical experiments (see e.g. [Körding *et al.* 2007, Vul & Rich 2010, Whiteley & Sahani 2008, Wozny *et al.* 2010]). Those and other studies provided some evidence that the brain has access to and computes with approximations or even the full probability distributions when doing inference. In the last decade the field of machine learning did provide theorists with many approximation techniques to reduce the complexity of computations that made this idea more plausible to be implemented in the brain (overviews in [Bishop 2006, MacKay 2003]). Despite the growing number of models proposed for probabilistic computations, a plausible mapping of those algorithms onto neuronal substrates is still not known. In contrast, there exists a number of experiments showing that for certain tasks results are contradicting some of the implications of the models [Brayanov & Smith 2010, Butler *et al.* 2011, Michel & Jacobs 2007, de Winkel *et al.* 2010]. Additionally, a number of important questions are not addressed by all of these theories and will require further research [Fiser *et al.* 2010, Jacobs & Kruschke 2011, Rothkopf *et al.* 2010, Triesch *et al.* 2010].

On the lowest level, that of the *implementation*, there exist some recent proposals on Bayesian inference with neurons (e.g. [Deneve 2008a, Ma *et al.* 2006, Soltani & Wang 2010]). Other models use a more bottom-up approach and try to match experimental data (e.g. [Anastasio & Patton 2003, Patton & Anastasio 2003, Rowland *et al.* 2007c, Ursino *et al.* 2008]). We will describe the most important of these in detail in section 2.2.4 of this introduction. The data to test or develop those theories mostly come from neurophysiology experiments on animals or to some degree also from neuroimaging studies.

Importantly, the recent findings of the inability of infants and young children to perform in accordance with predictions from Bayesian models [Gori *et al.* 2008, Nardini *et al.* 2008, Nardini *et al.* 2010, Neil *et al.* 2006], challenge many of the theories at the *algorithmic* and *implementational level.* These results point to a gradual development of inference computations. This thesis will propose a model that incorporates both the learning aspect and the near-optimal adult behaviour and how it can address those both *algorithmic* and *implementational level.*

To provide the necessary background knowledge, this Chapter first introduces Bayesian Theory, with a focus on cue integration, then provides an overview of experiments performed on various questions related to cue integration. It will also discuss existing computational models, and finally introduce reinforcement learning both from a theoretical and a biological side.

2.1 Bayesian Inference for Perception

The Bayesian view on perception states that there is a probabilistic relationship between perceived signals and the underlying states of the world that are to be inferred. Structural knowledge about causal relationships between different variables can be used to improve the outcomes. Bayes' Theorem formulates a way to compute a conditional probability distribution if only the inverse relation is known. This means knowing the generating process underlying observations along with some prior information, allows to infer the underlying causes based on the current stimuli. Mathematically that is:

$$p(X|Z) = \frac{p(Z|X)p(X)}{p(Z)}, \tag{2.1}$$

with Z being the observed variables, or cues, and X the quantities of interest. The benefit of this view over more classical interpretations is the accessibility of the full probability distribution instead of just a single estimate. The term on the left is usually termed posterior probability, the conditional term on the right likelihood, representing knowledge about the generation of stimuli. The second term in the numerator is referred to as the prior and could include knowledge from previous encounters with the given causes

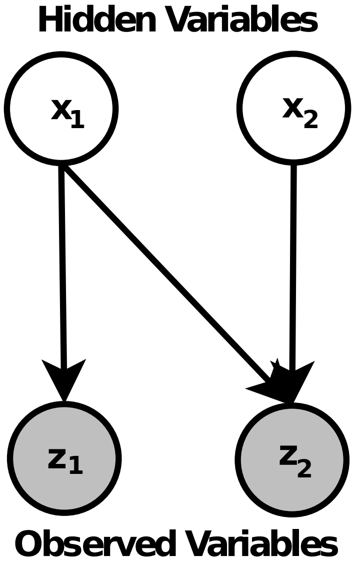

Figure 2.2. Graphical model Example for a graphical model showing the dependencies between two hidden (white) and two observed (gray) variables.

or represent an innate expectation. Finally the denominator takes care that the final result will sum to one. It will often be ignored if people are only interested in the maximum or if probability ratios are used.

A behavioural expression will usually be a single action and not the full posterior, therefore an additional computation has to be performed. The most common method for that is the maximum *a posteriori* estimate (MAP estimate) which simply uses the value at the peak of the distribution. Alternatives include sampling from the posterior or taking its mean. As will be discussed in section 2.3, for many tasks it is better to also include the potential reward of an estimate when selecting it.

The formulation of the posterior is supposed to include the structural knowledge present about the current environment m (to be correct one would write $p(X|Z, m)$). A common way to visualize such structure is shown in Fig. 2.2, where filled circles show observed variables, empty circles show hidden underlying causes (usually including the variable of interest) and connections between them show dependencies. For example, cues z_is that are independent of X will be excluded from the computation from the beginning. In the model shown in Fig. 2.2 for example, the observed variable z_1 does not provide information about the hidden variable x_2, and we can therefore write $p(x_2|z_1, z_2) = p(x_2|z_2)$. Additionally, if all cues in Z are independent one can factorize the likelihood term as

$$p(X|Z) = \frac{\prod_i p(z_i|X)p(X)}{p(Z)}, \tag{2.2}$$

which decreases the dimensionality of the computation enormously because we do not need the joint probability (of dimensionality $|Z|$) but only each cue's likelihood (of the dimensionality of z_i).

Which structure to use when doing inference is another important problem to be solved. When doing an experiment one always knows the generating model for the data explicitly. But human subjects may not have this information and therefore have to learn the task structure [Braun *et al.* 2010] or at least choose it from a number of predefined structures [Michel & Jacobs 2007]. Inference about the environmental model greatly increases the complexity of the computation, therefore some studies propose limitations to the set of possible models [Kemp & Tenenbaum 2008]. For each model from such a set one has to compute its probability given the data $p(m|Z)$, using Bayes rule this requires a prior over models $p(m)$ and the likelihood of the data given a model $p(Z|m)$. This model likelihood in turn is the data likelihood marginalized over all possible hidden states.

$$p(Z|m) = \int p(Z|X, m)p(X, m)dX. \tag{2.3}$$

To get to the full Bayesian posterior for the hidden variables $p(X|Z)$ one has to marginalize over models, or more intuitively weight the result under each model with that model's probability. This approach is called model averaging (MA).

$$p(X|Z) = \int p(X|Z, m)p(m|Z)dm \tag{2.4}$$

Model selection (MS) is slightly easier to compute, that is to pick the MAP estimate model and only perform inference on it instead of doing it on all possible models. In section 2.2.2 we will discuss psychophysical experiments that try to describe human behaviour with respect to this structural inference.

For cue integration one usually considers the setup, where multiple observables provide information about the same hidden variable (e.g. x_1 in Fig 2.2). Experimental studies expect the observed values to be noisy versions of the underlying variable. This noise is thought to come from internal and external sources, where the model setup often tries to make the external noise dominant since it can be controlled by the experimenter. Such noise will often be set to be additive, Gaussian with variance σ^2 and independent between the cues. The result of an integration of two such cues will also be Gaussian and have a smaller/equal variance than each of the single cues:

$$\sigma^2_{\mathrm{z}_1,\mathrm{z}_2} = \frac{\sigma^2_{\mathrm{z}_1}\sigma^2_{\mathrm{z}_2}}{\sigma^2_{\mathrm{z}_1} + \sigma^2_{\mathrm{z}_2}}, \tag{2.5}$$

in other words, an action based on this new distribution will be more reliable compared to one based on either cue alone. In the above case the mean of the final distribution will simply be a weighted average of the single cue means, where the weights are proportional to the inverse variances:

$$\mu_{\mathrm{z}_1,\mathrm{z}_2} = \frac{\frac{1}{\sigma^2_{\mathrm{z}_1}}}{\frac{1}{\sigma^2_{\mathrm{z}_1}} + \frac{1}{\sigma^2_{\mathrm{z}_2}}}\mu_{\mathrm{z}_1} + \frac{\frac{1}{\sigma^2_{\mathrm{z}_2}}}{\frac{1}{\sigma^2_{\mathrm{z}_1}} + \frac{1}{\sigma^2_{\mathrm{z}_2}}}\mu_{\mathrm{z}_2}. \tag{2.6}$$

Using this as an approximation to full Bayesian inference is very convenient, since it will be equal to the MAP estimate for simple cases like the one given above and also easy and fast to compute. Therefore many of the experimental studies we will introduce in the next section compare human performance to results of this kind of equation (e.g. [Ernst & Banks 2002, Johnston *et al.* 1993, Johnston *et al.* 1994]). Although this is working well in a controlled laboratory setting, the approximation might be much worse for natural problems with unknown structure, as we will show in section 4.3.

2.2 Cue Integration in the Brain

In this section I will provide an overview of experimental findings in tasks of cue integration and of the related model inference, paying special attention to the developmental aspects. It is split into a behavioural and a neurophysiology part in resemblance of the two categories of Marr that we want to address as stated above. Finally I will introduce existing models and highlight the open question that I want to answer in this thesis.

2.2.1 Behavioural Findings

Early studies of multisensory integration[1] already showed that human performance can be enhanced if subjects are provided with an additional stimulus in a second modality [Child & Wendt 1938,Thomas 1941, Todd 1912]. Generally there are two main ways in which the performance can be improved – additional cues can speed up the response (e.g. [Gielen *et al.* 1983, Hershenson 1962]), including the decrease of detection thresholds [Frens *et al.* 1995, Rach *et al.* 2011, Stein *et al.* 1989], or decrease the variability of responses (e.g. [Ernst & Banks 2002, Johnston *et al.* 1993, Johnston *et al.* 1994]), with a potential third way of removing biases and shape stimulus-response profiles in one cue by using a second established one (e.g. [Atkins *et al.* 2001, Bruns *et al.* 2011, Lackner 1973, Zaidel *et al.* 2011]).

Results from experiments with multisensory stimuli in terms of response facilitation can be overshadowed by other effects. If the response to only one of the stimuli is measured, a simple alerting (or response preparation) effect is often found, where the second stimulus is used as information about the point in time at which subjects should pay special attention [Diederich & Colonius 2008, Nickerson 1973]. The value of the signal itself does not provide information helpful for the task, therefore it can enhance reaction time but not acuity [Teder-Sälejärvi *et al.* 2005]. In contrast, if signals from all modalities carry relevant information beyond timing, one usually finds both improvements. But also with redundant information, the effect could just result from a stochastic process. This can be seen when considering the decision process of a single unimodal trial to be a stochastic race to threshold (also called drift diffusion model (DDM)) as is now commonly accepted (see reviews by [Ditterich 2010, Gold & Shadlen 2007]). The brain accumulates noisy information for or against a response over time, the response that reaches a certain threshold first is performed by the subject. If you would now have two independent races of this kind and choose the response of the one winning first, due to the stochasticity of the process you will see a decrease of the average reaction time, without any interaction between the two modalities. This formulation was called the "race model" [Raab 1962]. Modern studies will only call findings multisensory facilitation if the speed-up is stronger than what could be predicted by this race model [Miller 1982] (for examples see [Barutchu *et al.* 2009a, Hughes *et al.* 1994, Savazzi & Marzi 2002]). Although such multisensory facilitation was found for many task setups, studies comparing human reaction time with optimal predictions from Bayesian theories are still lacking. Such a Bayesian formulation would integrate the stochastic information from multiple cues at each time step and feed the result into a single race to threshold. Only very recently first results from a combined experimental-modeling study did show that humans are indeed producing reaction times close to the Bayesian predictions [Drugowitsch *et al.* 2010].

On the contrary when focusing on the response variability/acuity in multisensory tasks it is by now standard to compare the results with a Bayesian model (see reviews in [Ernst 2004, Kersten & Yuille 2003, Kersten *et al.* 2004, Knill & Richards 1996, Knill & Pouget 2004, Rothkopf *et al.* 2010]). Those studies have provided convincing evidence that humans combine sensory signals so as to reduce the uncertainty in their estimates used for responding. In many cases the response variability representing this uncertainty was close to the predictions from the optimal model. This is true for stimuli from across modalities such as in the judgment of the position of an object based on visual and auditory cues [Alais & Burr 2004, Battaglia *et al.* 2003, Binda *et al.* 2007], object size given visual and haptic cues [Ernst & Banks 2002] or information from vision and proprioception for trajectory discrimination [Reuschel *et al.* 2010, van Beers *et al.* 1996, van Beers *et al.* 1999]. Similarly, experiments have considered cues within the same modality as in inferring surface slant from stereo and texture cues [Hillis *et al.* 2004, Knill & Saunders 2003] or depth from texture and motion cues [Jacobs 1999]. Additionally priors seem to also be used similarly and are

[1] In most of this thesis we will use the terms multisensory, multimodal and cue integration equivalently to refer to the use of multiple information sources to solve a single task. Note however that in a strict sense, multisensory and multimodal integration are only referring to early sensory information (vision, audition, touch, proprioception, smell, taste), whereas cue integration is more general and also includes sources that are computed only within the brain (e.g. colour, shape, disparity, sound pitch, motion). As you will see in the rest of this introduction, the computational principles do not seem to differ between these classes and therefore the theoretical models are supposed to be general among them as well. Many experimental studies use the word multisensory whereas theoretical work often talks of cue integration.

optimally combined with each other and other cues as well [Mamassian & Landy 2001, Morgenstern *et al.* 2011], and can help to explain certain asymmetries in human and animal inference [Fischer & Pena 2011, Mamassian & Goutcher 2001, Weiss *et al.* 2002].

Most task setups are very similar in that they require subjects to do a two-alternative forced choice (2-AFC), determining if the second stimulus (called the "probe"), which is displayed varying in the dimension of interest, is bigger or smaller (or respectively left/right, farther/closer etc.) than a first stimulus (the "standard"). Response variability in unimodal trials is recorded first to get access to the reliabilities[2] of the single cues. These are used to define the optimal weights for Bayesian averaging (see section 2.1), when testing with multisensory stimuli. The weights used by the subjects are then accessed by showing them a standard with small mismatches between the values of the different cues, that are not noticed by the participants. Looking at the point of subjective equality (PSE), the value of the probe at which subjects choose either response with equal probability, one can determine how much the position of the standard was biased to the value of one or the other cue. It should be mentioned that in most of the studies only the average of the weights of all participants is close to the prediction of the Bayesian computations, individual subjects do not all show a perfect match. The general interpretation is that the brain might not use exact Bayesian methods but nevertheless uses the main principle of computation.

The procedure of finding the sensory weights is also used at different levels of external noise added to the stimuli to test the adaptability of the subjects' integration weights. It was shown that humans (and monkeys) can rapidly change the weights dependent on the current relative reliabilities of the cues [Fetsch *et al.* 2009, Triesch *et al.* 2002] (but also see [Zaidel *et al.* 2011]). In most of these studies such noise is thought to be directly encoded in a stimulus, when for example a visual stimulus is made less visible by reducing its contrast. A different approach uses artificial cues whose reliability is determined by statistical relationships to the hidden variable, they are e.g. only correlated with it in 50% of the trials. But even in this case it could be shown that humans can extract the correct cue weights with training [Atkins *et al.* 2001, Seitz *et al.* 2007] and sometimes even adapt very fast [Seydell *et al.* 2010, Triesch *et al.* 2002]. One possible mechanism of how weight updates can be calculated, despite the lack of direct feedback in many of those studies, could be the calibration of the changing cue through the more constant one [Atkins *et al.* 2003, Ernst *et al.* 2000, Ernst 2007]. Related to that it was also found that perceptual biases in one cue can be corrected for by another one. Which cue is adapted is determined again by the relative reliabilities. A classical example is the so called "ventriloquism aftereffect", where introducing a mismatch between auditory and visual stimulus position for a number of training trials leads to a later bias in the less reliable auditory modality in the direction of the visual training offset [Lewald 2002, Recanzone 1998, Wozny & Shams 2011] (see e.g. [Bruns *et al.* 2011, Burge *et al.* 2010] for similar effects in other modalities).

Developmental Findings

Although the interest in potential differences for multisensory perception in infants compared to adults did exist before [Lewkowicz & Turkewitz 1981, Meltzoff & Borton 1979, Spelke 1976], only recently studies are addressing potential developmental aspects in cue integration [Bahrick *et al.* 2002, Barutchu *et al.* 2009b, Lewkowicz 1996, Morrongiello *et al.* 1998a, Nardini *et al.* 2008, Needham 1999, Neil *et al.* 2006, Ross *et al.* 2011]. Those studies show a great variety of results, from 29 day old infants being able to detect congruences between auditory and visual signals [Meltzoff & Borton 1979, Morrongiello *et al.* 1998a] to 4 month olds only using a single cue to discriminate objects [Nardini *et al.* 2008, Needham 1999]. Multisensory facilitation of reaction times is immature for saccade tasks until 8 month [Neil *et al.* 2006], for more complex motor tasks until the age of 10 years [Barutchu *et al.* 2009a] The speed of learning on how to use multiple information sources seems to depend on the type of cue and maybe even the task.

Even more interestingly, when quantitatively testing the benefit of cue integration, people found that at least up to the age of 10 children are not as effective as adults [Barutchu *et al.* 2009b]. Very

[2] Reliability is used in this work as notation for the inverse variance of the noise distribution of a cue.

recently this was done in a principled way by comparing children's performances with that of Bayesian predictions [Gori *et al.* 2008,Nardini *et al.* 2008,Nardini *et al.* 2010]. Gori and colleagues [Gori *et al.* 2008] for example tested children between 5 and 10 years in two visuo-haptic discrimination tasks, for which adults show close-to-optimal integration. Participants had to do a 2-AFC for either the size of an object or its orientation in a plane. And whereas the older children showed adult-like integration, the younger group seemed to almost exclusively rely on the more reliable modality – haptics for the size, vision for the orientation task. The PSEs seemed to gradually evolve with age towards the Bayesian predictions (Fig. 2.3).

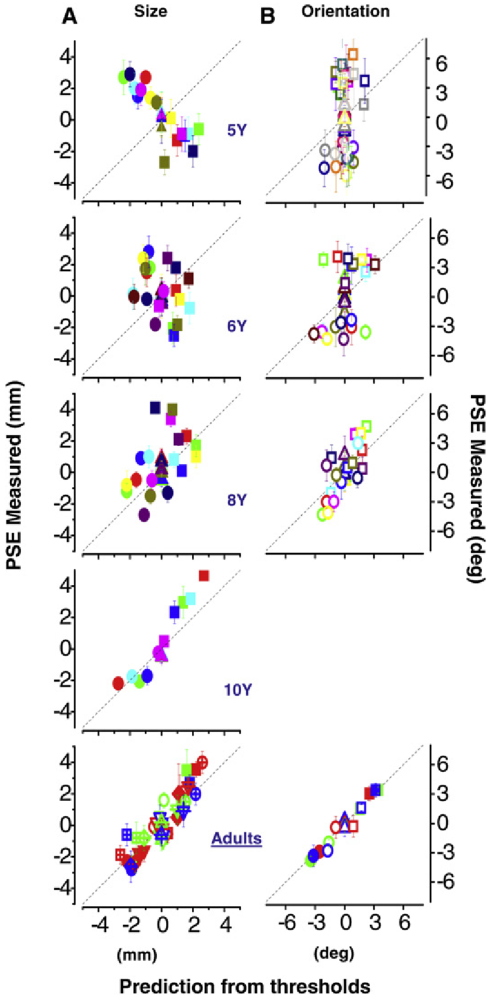

Figure 2.3. Development of cue integration. The PSEs for size (left) and orientation discrimination (right) of different age groups (from 5 years at the top to 10 years and adults at the bottom), plotted against the prediction from the Bayesian model. Each coloured point represents the true and predicted PSE of one subject. One can see a gradual development towards optimal integration. Figure reprinted from [Gori *et al.* 2008] (©2008, with permission from Elsevier).

The work of Nardini et al [Nardini *et al.* 2008] tested cue integration in children for a higher level task, namely navigation using self-motion and visual landmark cues. Both groups of 4-5 and 7-8 year olds did not integrate the two cues when available, but rather seemed to alternately rely on only one of them. Adults in contrast behaved in accordance with the prediction from optimal weighted averaging. In another study the same authors found similar results for children up to 10 years in a depth from texture and disparity task [Nardini *et al.* 2010].

All these studies show that, although the potential to use information from multiple cues seems to be present at least from early infancy, the ability to exploit the full potential benefit from cue integration develops over many years. A recent study [Putzar *et al.* 2007] points to an important role of experience as a driving force, speaking against a mere anatomical maturation as explanation. In experiments with children born with dense binocular cataracts, the authors show that visual deprivation during the first month of life impaired audio-visual integration even after complete recovery of sight (after treatment). Similar results were found for audio-visual speech recognition for children born deaf but using cochlear implants [Schorr *et al.* 2005]. Moreover it is worth mentioning that it was also found that early visual deprivation can lead to a worse-than-normal performance in other modalities for tasks where healthy subjects have highest reliability for visual estimates (shape [Held *et al.* 2011] or orientation [Gori *et al.* 2010]). These results support the idea of intermodal calibration as a developmental mechanism, where the better cue teaches the others [Gori *et al.* 2008, Gori *et al.* 2010] (but see also [Nardini *et al.* 2008]). This theory could be a first step in explaining the development of cue integration, but further theoretical discussion is still required. In Chapter 3 I will address this lack of theoretical investigation by proposing one possible mechanism of how this development can come about.

2.2.2 Causal Inference

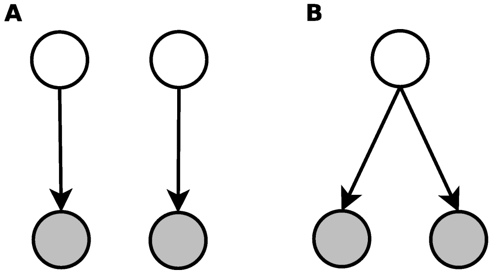

Figure 2.4. Generating models considered in causal inference. Two generating models with the same number of observables (cues) but one or two causes. In causal inference the task is to find out which of them produced the current input (with which probability). **A**: Two separate sources each produced one observation. **B**: One common source produced two observations.

The use of the Bayesian inference framework for cue integration in psychophysics has recently been extended to cases in which the observed sensory signals can be caused by one of two different scene configurations [Cheng *et al.* 2007, Körding *et al.* 2007, Rojas 2010, Sato *et al.* 2007]. This is interesting because humans obviously do not always integrate all signals from multiple modalities. In a natural environment the brain receives multiple signals in many modalities at each time point, most of them will not be of interest for the current task. It is important to determine which of them are relevant and then which may share the same source and could therefore be integrated to improve performance (Fig. 2.4B) or originate from different sources and should therefore be handled separately (Fig. 2.4A). The former might be solved by attentional mechanisms (which will not be discussed in this work, for a review see e.g. [Talsma *et al.* 2010]), the latter requires the selection of the underlying generating model[3].

[3] This problem also resembles a mechanism termed perceptual binding [Singer & Gray 1995, Treisman & Gelade 1980,

This second process has been termed causal inference [Körding *et al.* 2007, Shams & Beierholm 2010][4]. Computing the exact probability of a common cause for all possible sizes and combinations for groups of stimuli in a scene will usually be intractable. To still be able to compare human behaviour with the Bayesian predictions laboratory experiments have to reduce the complexity of the signals dramatically. Körding and colleagues [Körding *et al.* 2007] for example used a simple audio-visual orienting task with a single auditory stimulus z_a and a single visual stimulus z_v (see also Fig. 3.1 Top). By representing the uncertainty about the scene configurations, the Bayesian framework can be used to compute a posterior probability distribution over the two generating models. If only the two models mentioned above (common cause: $C = 1$ and two causes: $C = 2$) are considered, causal inference in this task boils down to:

$$p(C = 1|z_a, z_v) = \frac{p(C = 1) \int_x p(z_a, z_v|x)p(x)}{p(z_a, z_v)} \tag{2.7}$$

$$p(C = 2|z_a, z_v) = \frac{p(C = 2) \int_{x_a,x_v} p(z_a|x_a)p(z_v|x_v)p(x_a)p(x_v)}{p(z_a, z_v)}, \tag{2.8}$$

where one considers the two modalities to be independent, and generally only needs to compute one of the posteriors since $p(C = 2|z_a, z_v) = (1 - p(C = 1|z_a, z_v))$. If the model with common cause is correct there is only one true position x, whereas in the other case there exists an auditory and a visual position (x_a,x_v). To have better access to the underlying computations, Körding et al. asked their participants to report both the position of the auditory as well as the visual stimulus. Comparing these data with different theoretical models they found it best matched by the one using a probabilistic formulation. In a second experiment they explicitly asked the subjects to also report the number of causes in a given trial and again found close-to-optimal behaviour. In the first experiment the optimal behaviour includes marginalizing over the two possible scene configurations when computing the posterior of the auditory and visual positions. This computation is often called model averaging (MA). In the second experiment in contrast an optimal observer would have to choose the model with the higher posterior and then pick the positions according to that model. We will refer to this strategy as model selection (MS). Intermediate between these two strategies would be probability matching (PM), where the subject chooses one of the models but does so probabilistically based on the posteriors. In most experimental setups however, it is difficult to distinguish which of these strategies best matches human behaviour (see also the predictions for our toy example in Fig. 3.7) or whether humans use a single explicit strategy at all [Wozny *et al.* 2010].

The two most basic and also most effective information sources affecting the likelihood for a common cause $p(z_1, z_2|C = 1)$ are temporal and spatial distance between the observed signals. The influence of these two variables on subject's integration strategy was also validated experimentally [Gepshtein *et al.* 2005, Lewald & Guski 2003, Thomas 1941]. A famous example for MS using temporal synchrony as main clue is the ventriloquism effect – the voice and the motion of the lips of the puppet are perceived simultaneously and the brain therefore binds them together, and integration of the position estimates leads to the illusion that the voice originates from the puppet. The experiments used by [Körding *et al.* 2007] are inspired by this effect. Other groups tested the robustness of the perception of a single cause against spatial and temporal disparities directly and did find "windows of integration" along both dimensions [Lewald & Guski 2003, Slutsky & Recanzone 2001, Wallace *et al.* 2004b]. At least the temporal aspect seems to be plastic in adults: If people are exposed to audio-visual stimuli with a consistent

von der Malsburg 1995]. In the historical context "binding" was referring to the problem of having a coherent percept of an object despite of the separate representation of its features in early processing areas of the brain. But formulated in a probabilistic framework, the task is nothing else but inferring the probability of a common cause for a group of those features (which we call cues in this work) [Kersten *et al.* 2004].

[4] The term causal inference, along with a number of similar words like causal learning or causal reasoning was already used earlier to describe a more general process [Gopnik *et al.* 2004, Sobel *et al.* 2004, Tenenbaum *et al.* 2011]. It was referring to the reasoning about the causal structure between two or more variables. Described using a graphical model, this would be deciding which of the variables (circles) will be connected and in which direction. In our case inference is limited to comparing between two fixed causal structures.

temporal disparity, humans adjust their causal inference to better match this input structure (they shift the so-called point of subjective simultaneity) [Fujisaki *et al.* 2004, Harrar & Harris 2008].

Developmental Findings

There are few data about the development of causal inference abilities, and what is there is not using the comparison with an optimal model. But experiments did show at least that the temporal window of integration, the maximum time delay between two stimuli at which they are still integrated, changes with age [Lewkowicz 1996]. The size of this window decreases by almost a factor of 10 between early infancy and adulthood. Infants also seem to first rely on temporal synchrony for causal inference and only later additionally incorporate spatial disparity [Morrongiello *et al.* 1998b]. Generally infants seem to integrate cues much more frequently [Lewkowicz & Ghazanfar 2006, Pons *et al.* 2009] (although seemingly not as efficient, see subsection 2.2.1), which could mean that they do not use all available information for causal inference. This "perceptual narrowing" [Lewkowicz & Ghazanfar 2009], the extent of stimuli to be integrated decreases with age, also seems to be lost for some forms of autism disorder. It was shown that older autistic children still have the very large temporal integration windows found in young healthy infants [Kwakye *et al.* 2011].

2.2.3 Physiological Findings

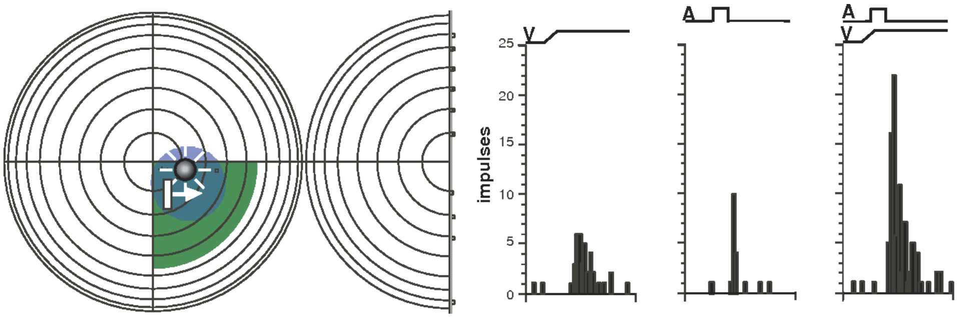

Figure 2.5. Audio-visual receptive field and firing rate of a neuron in cat superior colliculus. Left: Depiction of the auditory (green) and visual (blue) receptive field of an exemplary neuron. The RFs in the two modalities are largely overlapping. Receptive fields are plotted on a representation of visual and auditory space, with concentric circles at 10°spacing and the horizontal and vertical lines indicating 0°azimuth and 0°elevation, respectively. The half circle to the right represents the space behind the interaural axis. The two symbols represent the position of an exemplary visual and auditory stimulus. Right: Response profile of the same neuron to uni- and multisensory stimulation. The firing rate elicited by a multimodal stimulus is higher than the sum of the unimodal firing rates. Figure adapted from [Wallace & Stein 2007].

Early neurophysiological research did find neurons that can be activated by more than one sense in many different brain structures [Bruce *et al.* 1981, Junga *et al.* 1963, Meredith & Stein 1983, Murata *et al.* 1965, Wepsic 1966]. The most principled work comes from the superior colliculus (SC), a subcortical structure that has large numbers of multisensory neurons [Meredith & Stein 1986, Wallace *et al.* 1996] that are spatially aligned and organized topographically [Stein *et al.* 1993]. The SC also has a direct behavioural connection through its involvement in saccade generation [Burnett *et al.* 2007, Sparks & Mays 1990], so it is well suited to study the connection between multisensory neurons and behaviour [Leo *et al.* 2008]. One of the important principles of operation in multisensory neurons is the nonlinearity in the combination of inputs. When comparing the firing rate of a neuron in response to a multisensory input with the sum of the firing rates in response to either unisensory input (e.g. Fig 2.5), one can find both superadditivity – a response higher than the sum, and subadditivity – a response lower than that sum [Avillac *et al.* 2007, Meredith & Stein 1983, Meredith & Stein 1986, Stanford *et al.* 2005], as well as a suppression, where inputs from a second modality lower the response to the first one [Kadunce *et al.* 1997, Wallace *et al.* 1996].

The unisensory receptive field (RF) of the great majority of SC neurons are overlapping in their spatial profile [Avillac *et al.* 2007, Wallace *et al.* 1992, Wallace *et al.* 1996] (Fig 2.5; But also see [Slee & Young 2011]). Additionally, suppression is often found if one of the stimuli is far away from the other (and outside the RF) [Kadunce *et al.* 1997, Wallace *et al.* 1996]. This so called spatial principle could be seen as part of an implementation of causal inference – two stimuli from the same position are very likely to come from a common source (see subsection 2.2.2). Similarly a temporal principle, that two stimuli that appear to be close in time are more likely to share the same source, can be found in those neurons as well [Meredith *et al.* 1987]. Again stimuli too far apart in time seem to even suppress the neuronal response. A second factor determining the type of response is the strength of the incoming stimuli. If both signals are strong one often finds subadditivity, if both are weak the response is stronger than their sum. This phenomenon has been named the principle of inverse effectiveness [Meredith & Stein 1983, Stanford *et al.* 2005, Wallace *et al.* 1996].

Such non-linear enhancement if combined with the idea of a DDM for decision making could explain the enhanced reaction times to multisensory stimuli [Rowland *et al.* 2007a]. There is also evidence that the receptive field structure of these multisensory neurons is in qualitative accordance with Bayesian predictions [Rowland *et al.* 2007b].

Developmental Findings

There seems to be a strong developmental influence on multisensory integration at the physiological level as well (see e.g. review by [Wallace 2004]). In cat SC there are no multisensory neurons found until 10 days after birth, instead it seems that first unisensory neurons develop of which afterwards many get responsive to a second modality [Wallace & Stein 1997]. It still takes 3 to 4 month after that onset before the response profiles of those multisensory neurons are similar to those found in adults [Wallace & Stein 2000]. Newborn monkeys already have multisensory neurons in their SC, but as in cats their RFs are much larger and less aligned than those of adults. And if stimulated by a multisensory signal within their RFs, the response of those neurons is not different from stimulation with the more effective single cue signal, they seem to not yet integrate the two inputs [Wallace & Stein 2001]. Similar results can be found in a cortical multisensory area that sends inputs to SC, the anterior ectosylvian sulcus (AES) [Wallace *et al.* 2006], and these two areas seem to be developmentally coupled [Wallace & Stein 2000]. Ablating both these areas impaired adult cats from integrating auditory and visual signals, removing only one of them instead seems to be compensated for by the brain during development [Jiang *et al.* 2007]. Rearing kittens in darkness does not prevent the occurrence of neurons responsive to more than one sense during development, but it does impair these neurons from integrating those inputs in a way that would exceed unisensory firing rates [Wallace *et al.* 2004a]. The same is true if kittens get normal visual and auditory inputs during their first weeks of life, but never experience combined audio-visual signals [Yu *et al.* 2010]. These studies clearly speak against a purely anatomical explanation of the development of cue integration.

Maybe even more interestingly, physiological studies also found early evidence for developmental plasticity in causal inference-like processes. The spatial alignment of the different sensory RFs, which as was said before can be seen as implementing spatial distance between two signals as an important clue to causal inference, was shown to be sensitive to early postnatal experience. Cats were raised in an artificial environment, in which auditory and visual signals were always shown at the same time but at systematically shifted spatial positions [Wallace & Stein 2007]. Subsequent behavioral tests revealed that the animals did not integrate multisensory stimuli from a common location, as seen in animals raised in natural environments, but instead integrated only signals with the distinct spatial separation present in the artificial environment. Recording from SC neurons also showed a misalignment in the unisensory RFs (Fig. 2.6, compare with normal RFs in Fig 2.5). Similar findings were also made earlier in owls [Knudsen & Brainard 1991], where raising animals with prisms shifted the auditory RF center of multisensory neurons in the direction of the prism distortion. There exist also behavioural experiments altering the temporal integration window. Quails that had no prenatal exposure to temporally synchronous audio-visual signals later failed to react to those stimuli (with no impairment in unisensory reaction), contrary to normally developing animals [Jaime & Lickliter 2006].

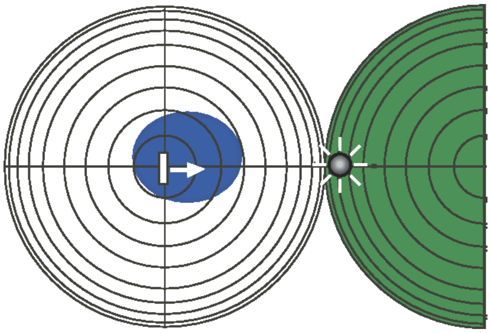

Figure 2.6. Misaligned audio-visual receptive field of a neuron in cat superior colliculus. Depiction of the auditory (green) and visual (blue) receptive field of an exemplary neuron (similar to Fig. 2.5) after raising kittens in an environment where auditory and visual signals always had a fixed non-zero disparity. The centers of the RFs in the two modalities are shifted to each other. Figure adapted from [Wallace & Stein 2007].

2.2.4 Models of Cue Integration

As mentioned at the beginning of this chapter, most scientist agree that the general purpose of perception is to infer the environmental variables that cause the sensory input. Following this line, the *computational level* interpretation of cue integration is simply to improve this inference process and lower the impact of internal and external noise of the inputs [Bülthoff & Yuille 1991, Knill & Richards 1996]. I also showed in section 2.2.1 that for humans the outcome of such inference processes is close to the prediction from the Bayesian theory. Full numerical Bayesian inference is not a very good theory at the *algorithmic level* though, since in requires a huge number of operations for natural inputs. One popular approximation proposed is the use of weighted averaging [Jacobs 1995, Johnston *et al.* 1993, Johnston *et al.* 1994, Landy *et al.* 1995, Taylor 1962] (see also sec 2.1). To use it, the brain would simply have to represent means and variances of the underlying probability distributions. The approximation is also known to be close to the true solution for most distributions with only a single maximum, which indeed seems to be the case for the likelihoods of many basic variables given a cue [Stocker & Simoncelli 2006]. For more complex setups, though, this assumption need not to be true. Consider for example color as a cue for object recognition – a certain perceived color will assign high probabilities to a large number of objects that do not necessarily have to be next to each other in some object dimension based for example on similarity in utilization. The same is true for some prior distributions (e.g. the multi-peaked natural orientation prior [Coppola *et al.* 1998, Rothkopf *et al.* 2009]). A second implicit assumption for this approximation is the independence between the noise of different cues, which allows each cue's likelihood term to be factorized in the Bayes formula (see eq (2.2)). Depending on the cues this might not always be the case [Elliott *et al.* 2009, Karaoguz *et al.* 2011, Oruç *et al.* 2003].

A model by Triesch and von der Malsburg [Triesch & von der Malsburg 2001] used weighted averaging in combination with online reweighing of the cues to adapt to changing environments. Each of their cues computed a spatial map of "saliency" values stating how similar a certain point in space is to the target object. These maps are then summed together, weighted by a reliability term. The reliability term was not explicitly the noise variance of a cue but rather represented a running average of the agreement of that cue with the integration result. Therefore the model did not have to know the variances in advance and it could quickly reweigh a cue that broke down, like for example a motion cue if the object of interest stops moving.

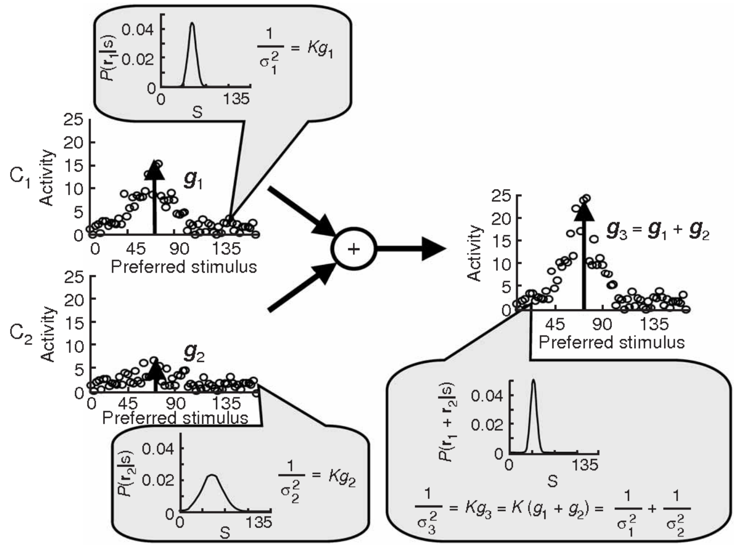

Figure 2.7. Probabilistic population codes for cue integration. The left side shows the activity of two populations of neurons, each representing one cue's estimate of a common hidden variable. Each population encodes the variance of the estimate through its gain g (K is a constant that depends on the tuning curves of the single neurons). Adding these two population activities, neuron by neuron (spatially aligned), will lead to a gain representing a variance equal to the prediction from Bayesian averaging (eq (2.5)). The maximum of the new population activity will be according to eq 2.6. Figure reprinted with permission from Macmillan Publishers Ltd: Nature Neuroscience ([Ma *et al.* 2006] ©2006).

One model porting the idea of weighted averaging to the *implementational level*, makes use of the stochasticity of neuronal spiking to explain cue integration using biologically plausible mechanisms [Beck *et al.* 2008,Ma *et al.* 2006,Ma & Pouget 2008]. In the so called probabilistic population code (PPC) [Sahani & Dayan 2003, Zemel *et al.* 1998] a probability distribution is encoded by the activity of a population of neurons. Each neuron fires probabilistically with a certain rate r given a stimulus s. The likelihood distribution $p(r|s)$ is determined by a Gaussian RF together with some independent Poisson noise. The posterior probability $p(s|R)$ will be the product of all the likelihoods and converges to a Gaussian for large numbers of neurons. The mean of this distribution will be close to the firing peak of the population and the variance will be inversely related to the height of this peek. If the response of each pair of spatially aligned neurons from two such populations, representing two cues, is added together, the resulting responses of the "multisensory" population will encode a mean according to eq (2.6) and variance according to eq (2.5) (Fig 2.7). In this model the ability to do optimal integration is inherent in the properties of the neurons, namely in their Poissonian stochasticity and their Gaussian tuning curves. This is on the one hand an elegant and efficient way of explaining how the brain is doing near-optimal inference, but on the other hand could only explain a developmental switch from no integration to full inference without being able to reproduce the intermediate behaviours found in experiments [Nardini *et al.* 2008, Gori *et al.* 2008].

Deneve developed a different model in which integration is done by single neurons [Deneve 2008b, Deneve 2008a] (see also a similar but less detailed proposal by [Anastasio *et al.* 2000]). The membrane potential of a neuron is thought to represent the posterior (or log-odds) ratio for a certain cause being present in the inputs. One benefit of this model is that it also uses the temporal relationships between spikes for coding. Additionally an electrophysiologial study in monkey LIP neurons showed a behaviour that resembled a log-probability ratio as well [Yang & Shadlen 2007], which can also be related to the aforementioned DDMs.

Another alternative proposal is that the brain samples from the underlying probability distributions without having to explicitly represent them [Mozer *et al.* 2008, Moreno-Bote *et al.* 2011, Vul *et al.* 2009, Vul & Rich 2010]. There exist many sampling algorithms from the machine learning field and for example Markov Chain Monte Carlo [Gershman *et al.* 2010, Hoyer & Hyvärinen 2003] and importance sampling [Shi & Griffiths 2010] were already proposed to be implemented by the brain. But both these models also share the implementation of Bayesian mechanisms as inherent features of the neurons, and can therefore not contribute to an explanation of the developmental aspects that we want to focus on.

A bottom-up modeling approach is used by groups trying to reproduce features found in neurophysiological experiments on multisensory integration [Anastasio & Patton 2003, Magosso *et al.* 2008, Ohshiro *et al.* 2011, Patton & Anastasio 2003, Rowland *et al.* 2007c, Ursino *et al.* 2008]. Those models try to mostly reproduce qualitative features of biological experiments, without testing them explicitly against optimal predictions. Many ideas present in these studies were also used in biologically inspired algorithms for more robotic applications [Monaci *et al.* 2009, Rucci *et al.* 2000, Schauer & Gross 2003, Wysoski *et al.* 2010].

The group of Barry Stein which produced many of the important neurophysiological work on multisensory integration developed a model of the dynamics of a single SC neuron [Rowland *et al.* 2007c]. They used multiple "dendritic compartments" within a single neuron to first integrate multisensory stimuli representing primary sensory areas and AES separately. The integration uses a delayed enhancement mechanism inspired by the two main types of receptors in biological synapses (fast AMPA receptors (AMPAR) and slower NMDA receptors (NMDAR)). The final integration step at the "cell body" is controlled by an additional inhibitory neuron also receiving inputs from those areas. The model reproduced many of the findings from their own recordings, like multisensory enhancement, inverse effectiveness and NMDAR dependence of the enhancement (from [Binns & Salt 1996]). Unfortunately, the critique that all these findings come from fixed structural features of the neuron and can therefore not explain the gradual development can also be applied here.

Anastasio and Patton [Anastasio & Patton 2003, Patton & Anastasio 2003] built an artificial neural network (ANN) using a synapse structure, where each of the main connections from a single modality is modulated by additional inputs from all unisensory areas. Afterwards they train the main and modulatory connection in two separate training stages with first a Hebbian and then an Anti-Hebbian learning rule. This setup is able to nicely reproduce two findings from recordings in cat SC, namely non-linear multisensory enhancement and the presence of both uni- and multisensory neurons in the same area. Additionally it is also constructed in a way to show a strong decrease in enhancement when turning off the modulatory connections (from AES) in accordance with experimental findings [Jiang *et al.* 2007]. Many parameters of this model are explicitly constructed to produce the desired results, e.g. map alignment, separate training steps and rules, which make it less interesting in terms of developmental considerations.

A similar approach was proposed by Ursino, Cuppini, Magosso and colleagues [Cuppini *et al.* 2010, Magosso *et al.* 2008, Ursino *et al.* 2008], but they focused on explaining effects on a network instead of a single neuron level. They are using two unisensory networks with fixed lateral connections that project to and receive projections from a multimodal network of the same structure. Each unisensory neuron has a RF centered on a region in space and is connected to the multisensory neuron representing the same position. Through the "Mexican hat" structure of the lateral connections (excitations for close-by neighbours, inhibition for more distant units), this model can well predict the mutual suppression of disparate stimuli. The synaptic weights are set to produce multisensory enhancement and to some point also reaction time facilitation. Again, despite these successful findings, it can only provide a prediction for the final structure of the implementation, since it does not use any plasticity mechanisms.

Very recently the same group tried to address this issue [Cuppini *et al.* 2011a, Cuppini *et al.* 2011b] by extending their model by a number of additional input structures and working with synaptic plasticity. They first showed that this model can reproduce their previous findings, and then set some of the synapses to zero to produce a "newborn" state were integration could not happen. These synapses were afterwards trained with uni- and multisensory inputs using a modified Hebbian learning rule with normalization. After learning the network was able to regain the desired features mentioned above.

Rucci [Rucci *et al.* 1997] and later others [Huo & Murray 2009, Mysore & Quartz 2005] developed models that tried to explain the spatial alignment of the unisensory RFs within a multisensory neuron. These groups were inspired by the prism experiments in barn owls (see section 2.2.3) and used variants of Hebbian learning rules to reproduce findings from those experiments. Rucci and colleagues used the fact that multisensory signals that occur at the same time tend to also come from the same location. The Hebbian rule enhances synapses with temporally correlated activity and by that promotes those that come from the same position in both unisensory maps. If a prism changes the relation of space and time disparity the resulting final maps will also show the shift. Interestingly these models use a reward signal depending on the success of foveation of a stimulus to only reinforce those multisensory signals that really come from the same object, an approach similar to what we use in our model. Nevertheless we think the work presented in this thesis goes well beyond their model by explicitly addressing the optimality of the causal inference, as well as taking a much deeper look at cue integration as the main reason for this development.

The last model to be mentioned tries to concentrate on explaining temporal facilitation [Colonius & Diederich 2004, Diederich & Colonius 2008]. It proposes that the unisensory areas each take a decision on themselves and, based on a relaxed race model, a second multisensory stage will only integrate if those races terminate within a certain time window. Since the model always waits until the end of this window, multisensory facilitation of reaction time comes from a delay in unisensory processing. This model does not try to explain the accuracy aspect nor any similarity to Bayesian predictions.

All the existing models tend to either focus only on explicitly reproducing a few biological effects or carry most of the desired properties already in their structure. Specifically, there is yet no model that can explain both a developmental progress as well as final near-optimal behaviour. In this thesis I will try to fill this gap with a new model based on reward-mediated learning.

2.3 Reinforcement Learning

How do animals learn from their experiences? If one agrees that most behaviour has the goal of being beneficial to the animal (e.g. in terms of survival), it makes sense to take a closer look at the interactions between an animal's own actions and changes in the environment. If we also have a way of measuring the subjective quality of an environment (that is the value of the current instance for the individuum), it will be possible to simply perform those actions more often that led to an improvement of this quality. Observing the outcome of an action can also teach you about causes and effects in general. One benefit is that we do not need an explicit teacher that provides us with the correct answer, but only a couple of sensors signaling changes in the environment. This type of learning is called reinforcement learning

(RL) [Sutton & Barto 1998]. A "reinforcement" is a measure of the current quality of the environment, it can be an explicit reward, e.g. food, or a more abstract signal like the happiness of finding your keys in your bag. RL is not only interesting for explaining biological learning processes but can also be helpful when constructing algorithms that have to be able to perform in unknown environments. In the following I will give a short overview of both theoretical aspects and evidence for a biological implementation of RL.

2.3.1 Theory

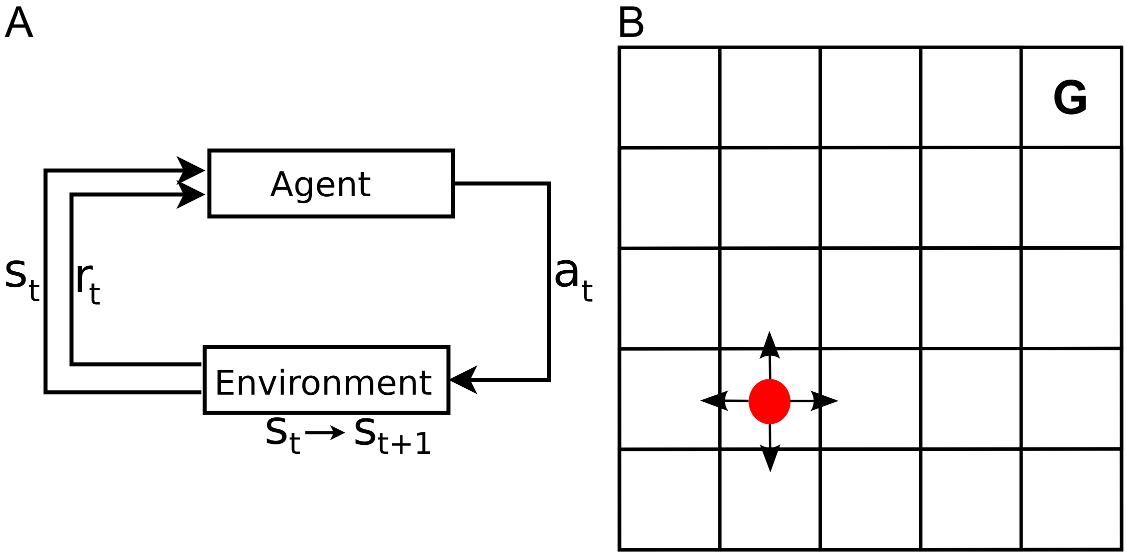

Figure 2.8. Reinforcement Learning. A: Schematic depiction of relations of variables in reinforcement learning. An agent perceives the state of the world s_t and the current reward r_t with its sensors and performs an action a_t; based on that the environmental dynamics change the state to s_{t+1}. **B**: A computational agent (red) has to move in a grid world starting in S, and receives a positive reward if reaching the goal state G.

The theory of reinforcement learning uses a very basic idea of environmental dynamics, namely a mapping $p(s_{t+1}|s_t, a)$ from a state s_t to state s_{t+1} which is mediated by an action a. A computational "agent" gets sensory information about its current state and chooses an action based on a policy π, a (probabilistic) mapping from s_t to a. The agent will receive a reward r_t (positive or negative) if its action gets it into one of a subset of states, based on the reward function $r(s_t)$. Figure 2.8A shows a schematic drawing of the process.

A common exemplary problem is the "grid world" shown in Fig. 2.8B, where the state is simply the current position. The agent has a starting position S and four actions move it (probabilistically) in the four principal directions. If the agent reaches the goal state G, it receives a positive reward. The goal of learning is to change the policy in a way that maximizes the reward that the agent receives.

In its most simple form this means increasing $\pi(a|s_t)$ if the next state s_{t+1} holds a positive reward, decreasing it if its negative. However such learning only works for states from which one can get into a rewarded state with a single action. To be able to solve more complex problems the agent has to compute a "value function" V. This function predicts how much accumulated reward can be expected in the long run starting in a given state s_t if using a policy π.

$$V^\pi(s_t) = E^\pi(R|s_t) = E^\pi(\sum_{i=0}^{\infty}\gamma^i r_{t+i+1}|s_t) \tag{2.9}$$

A discount rate γ $(0 < \gamma < 1)$ that decreases reward values with distance in time is often used, where a reward in the future is worth less than the same reward in the next time step. This is not only intuitive, but also enforces the agent to find the shortest path to the rewards. We can rewrite the above equation to arrive at a recurrent formulation called the *Bellman equation* that is using the current reward (r_{t+1}) and the current value function:

$$\begin{aligned} V^\pi(s_t) &= E^\pi(\sum_{i=0}^{\infty}\gamma^i r_{t+i+1}|s_t) \\ &= E^\pi((r_{t+1} + \gamma\sum_{i=0}^{\infty}\gamma^i r_{t+i+2})|s_t) \\ &= E^\pi((r_{t+1} + \gamma V^\pi(s_{t+1}))|s_t) \\ &= \sum_a \pi(s_t, a)\sum_{s_{t+1}} p(s_{t+1}|s_t, a)(r(s_{t+1}) + \gamma V^\pi(s_{t+1})) \end{aligned} \tag{2.10}$$

The next state will usually also depend on the action that is chosen, therefore it makes sense to also compute another value function $Q^\pi(s_t, a_t)$, that is defined similarly but depends on both current state and action (called Q-function or Q-value function). For small state spaces (state-action spaces) it is possible to simply store the expected reward values in a big table. However, as that space grows one usually has to start to approximate the value function instead. One example for such a function approximation method, artificial neural network (ANN)s [Sutton & Barto 1998], will be used for the work presented in Chapter 3.

To get V^π directly the agent would have to know the environmental dynamics and the reward function. Agents that have an explicit representation of the dynamics are called "model-based reinforcement learner". Those agents could either know those dynamics or learn the model of the environment from observations in parallel to the optimization of reward. Model-based methods are fast in learning the value function, because updating one state after receiving a reward will affect all previous state(-action) values causally connected to it. This happens because at each state the agent computes eq (2.10). The down side of this is that it makes deciding for an action computationally slow. The use of potential future states for a current decision can be seen as planing behaviour.

If the agent does not have access to the dynamics, i.e. it acts "model-free", it can use Monte Carlo sampling, where the agent samples an episode that ends in a terminating state and then, based on the overall reward, updates the predictions of all states that were visited during that episode. Alternatively, online updating of the value function immediately after moving to the next state s_{t+1} can be done using

$$V_{\text{new}}(s_t) = V_{\text{old}}(s_t) - \varepsilon(V_{\text{old}}(s_{t+1}) + r(s_t) - V_{\text{old}}(s_t)), \tag{2.11}$$

where ε is a learning rate parameter. This is called "temporal difference (TD) learning" because it uses predictions from two consecutive time steps. For a fixed policy, TD learning was shown to converge to the true V^π [Dayan 1992, Jaakkola *et al.* 1994, Sutton 1988]. Model-free RL can be seen as slow in learning, because it only updates values of states it visits, but fast in deciding, because it only has to check the values for the current state.

How are value function and policy related? The best policy π^* is one that has maximal $V(s)$ for all states s. Or rather if we know the optimal value function V^*/Q^*(according to eqs (2.12/2.13)), the best policy chooses the action that gets us to a next state that has the highest predicted $V^*(s_{t+1})$ (the action that predicts highest reward respectively for $Q^*(s_t,a))$ (called a "greedy policy"). V^* is defined by the *Bellman optimality equation* as equal to the expected return of the best action from the current state:

$$\begin{aligned} V^*(s_t) &= \max_a(Q^{\pi^*}(s_t,a)) \\ &= \max_a(E^{\pi^*}(R|s_t,a)) \\ &= \max_a(E^{\pi^*}(\sum_{i=0}^{\infty}\gamma^i r_{t+i+1}|s_t,a))) \\ &= \max_a(E^{\pi^*}(r_{t+1}+\gamma\sum_{i=0}^{\infty}\gamma^i r_{t+i+2}|s_t,a))) \\ &= \max_a(E^{\pi^*}(r_{t+1}+\gamma V^*(s_{t+1})|s_t,a))) \\ &= \max_a\sum_{s_{t+1}} p(s_{t+1}|s_t,a)(r(s_{t+1})+\gamma V^{\pi^*}(s_{t+1})) \end{aligned} \tag{2.12}$$

Accordingly, the optimal Q-function is:

$$Q^*(s_t,a)=\sum_{s_{t+1}} p(s_{t+1}|s_t,a)(r(s_{t+1})+\gamma\max_{a'}Q^*(s_{t+1},a')) \tag{2.13}$$

To learn the policy online, while we also learn the value function, we have to focus on state-action values:

$$Q_{\text{new}}(s_t,a_t)=Q_{\text{old}}(s_t,a_t)-\varepsilon(Q_{\text{old}}(s_{t+1},a_{t+1})+r(s_t)-Q_{\text{old}}(s_t,a_t)). \tag{2.14}$$

Because of the five variables used this update rule is called SARSA (s_t,a_t,r_{t+1},s_{t+1},a_{t+1}). In a state s_t the agent will choose the action a_t that maximizes $Q(s_t,a_t)$, observe the following reward and state and use this together with the next prediction to update $Q(s_t,a_t)$.

Algorithms like SARSA that use the action determined by the current policy π to update the Q-value function (like eq (2.14)) are called on-policy [Sutton & Barto 1998]. There is also an alternative strategy called off-policy, were the updates are according to the best possible action and therefore do not depend on π. An important example for this would be Q-learning which uses

$$Q_{\text{new}}(s_t,a_t)=Q_{\text{old}}(s_t,a_t)-\varepsilon(\max_a(Q_{\text{old}}(s_{t+1},a))+r(s_t)-Q_{\text{old}}(s_t,a_t)) \tag{2.15}$$

as update rule. Q-learning was shown to converge to Q^* with probability 1 in the limit of infinite samples of each state [Watkins & Dayan 1992].

Both types of algorithms require each state (state-action-pair) to be sampled a high enough number of times for Q to approach the true solution. The optimal behaviour thus would be to randomly sample the state-action space to produce good statistics for the value estimation. In contrast if we also want to maximize reward during runtime, a pure greedy policy could guarantee that for Q^*. But as long as we only have an approximation of it, this strategy could easily run into local optima and not ever visit certain states. This dilemma in the general case is called the exploration-exploitation trade-off. There is no optimal solution to this dilemma, so most people use stochastic policies, where for example in a small fraction ϵ of cases the agent will choose a random action and the rest of the time act greedily (ϵ-greedy policy). ϵ-greedy is among the learning policies for which SARSA could be proven to converge to Q^* with probability 1 [Singh *et al.* 2000].

Instead of directly deriving the policy from the value function, these two can be computed separately. Such an architecture is called actor-critic RL, the actor learns the best action for a given state and the critic learns the value function. Based on the temporal difference error of the critic both functions are updated. This is also an on-policy algorithm. Actor-critic methods are beneficial for cases with large (or continuous) action-spaces, because they do not necessarily have to search the full action space, since the policy is explicitly stored.

2.3.2 Biological Evidence

Reward-dependent learning of animals has been the topic of scientific research for the last century [Pavlov 1926, Thorndike 1911]. Most famous are Pavlov's experiments on classical conditioning, where a dog learned to expect a primary reward (food) whenever hearing an in itself unrewarding conditioned stimulus (bell). In terms of RL theory the dog learned to value the state "bell ringing" almost as much as the food which is always following, because the bell meant that he got closer to the reward. In the operant (or instrumental) conditioning paradigm [Thorndike 1911] a reinforcement signal is used to enhance or decrease certain active behavioural responses, like the pressing of a lever. There is ample data demonstrating reward dependent learning even for basic behaviours like orienting movements [Hikosaka *et al.* 2006, Platt & Glimcher 1999, Takikawa *et al.* 2002, Schultz 2000]. These principles also work for more distant reward, that is for example using a light to predict the bell that predicts the food.

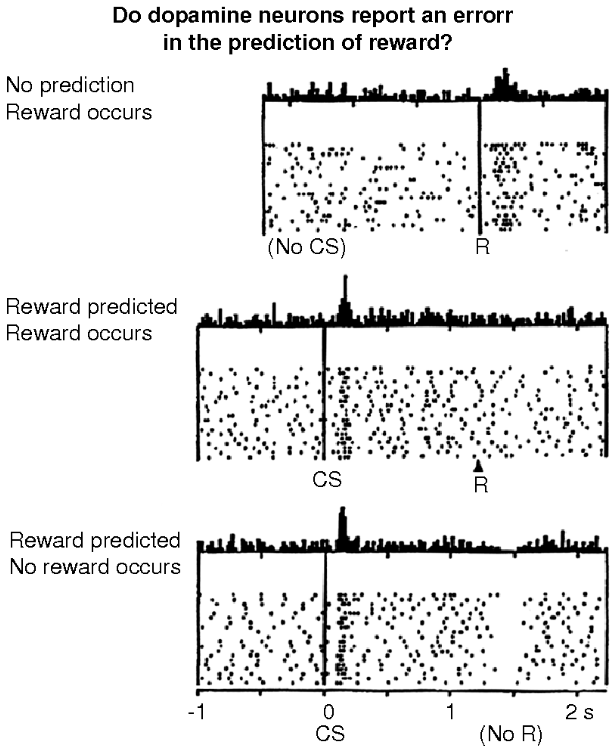

Figure 2.9. Dopamine neurons signaling reward prediction error. Rasterplots of the prediction error encoding responses of a dopaminergic neuron in VTA for a classical conditioning setup. **Top**: The dopaminergic neuron fires when a reward is given. **Middle**: After learning the firing rate increases when the conditioned stimulus is shown, at the time of reward the neuron fires at baseline. **Bottom**: If after showing the conditioned stimulus the reward is omitted, the neuron decreases firing below baseline. From [Schultz *et al.* 1997], reprinted with permission from AAAS.

In the last 20 years, researchers have also found neuronal activity correlated to certain variables used in the RL models introduced in the last section [Daw & Doya 2006, Rushworth & Behrens 2008, Samejima & Doya 2007, Yamada *et al.* 2011]. Most prominent was the finding that dopamine (DA) neurons in the ventral tegmental area (VTA) seem to encode a reward prediction error [Caplin *et al.* 2010, Montague *et al.* 1996, Schultz *et al.* 1997]. Before that, it was already shown that DA seems to act as a rewarding signal in the brain, enhancing behaviour that increases the DA dosage (e.g. [Phillips *et al.* 1976]) or that one can increase learning with DA increasing drugs [Pessiglione *et al.* 2006]. Midbrain DA releasing neurons increase their firing rate when the animal receives a drop of juice [Schultz 1986]. Schultz and colleagues [Schultz *et al.* 1997] could show that this increase in firing rate can be shifted towards the time when a conditioned stimulus (CS) was shown. The reward itself did not elicit any changes anymore in this setup (see Fig. 2.9). But if the reward that is predicted by the CS is omitted the VTA neuron will decrease its firing rate below the baseline firing. This can be interpreted as predicting a reward at the CS and if the reward appears as predicted no extra signal has to be submitted, otherwise firing decreases to show that the reward was worse than expected. Later additional experiments also showed that this signal also includes probabilistic predictions [Fiorillo *et al.* 2003, Morris *et al.* 2006] and even conditioned probabilities [Nakahara *et al.* 2004]. Human functional magnetic resonance imaging (fMRI) studies also showed correlations of various brain areas with RL variables like Q-values and prediction errors [Berns *et al.* 2001, D'Ardenne *et al.* 2008, Haruno & Kawato 2006, McClure *et al.* 2003, Pessiglione *et al.* 2008, Schönberg *et al.* 2007].

Besides the research on the encoding of certain RL variables by dopaminergic neurons in the brain, it is also interesting to look at how learning using the DA signal is done at downstream neurons. In cortex and other areas it was shown that DA modulates neural plasticity mechanisms (see e.g. reviews in [Calabresi *et al.* 2007, Jay 2003, Pawlak *et al.* 2010, Wickens 2009]) and that it is necessary for various learning tasks like motor learning [Molina-Luna *et al.* 2009]. In auditory cortex A1 for example, pairing a certain pure tone with reward leads to a sharpening of the tuning curves in the respective neurons and also selectively enhances signal transmission from those neurons to secondary auditory areas [Bao *et al.* 2001]. Activating DA receptors in the hippocampus lowers the threshold for the induction of standard synaptic plasticity like long term potentiation (LTP) and long term depression (LTD) [Lemon & Manahan-Vaughan 2006]. In other areas DA seems to be required to allow synaptic plasticity at all [Pawlak & Kerr 2008, Reynolds *et al.* 2001].

All these findings show the relevance of RL for learning in many different tasks and environments both on the behavioral as well as the computational level [Glimcher 2011]. It is thus interesting to consider it as one potential driving force for the development of cue integration and causal inference.

2.3.3 Multiple Controller Hypothesis

As stated in the theory section, in RL one can differentiate between model-based and model-free learning methods. In the past many groups tried to show that one or the other is implemented by the brain to control behaviour.

There is behavioural evidence for a model based RL system in the brain [Adams & Dickinson 1981, Balleine & Dickinson 1998, Colwill & Rescorla 1985]. A typical behavioural experiment used in this work is reinforcer devaluation [Rozeboom 1958, Tolman & Gleitman 1949]: The subject is trained in a multi-step task receiving a positive reward when reaching the desired final stage. After successful training the reward is devaluated, for example by adding poison to a food reward, and shown separately (this is inspired e.g. by the changing value of food reward based on hunger level). A model-free learner would continue to approach the "rewarded" state, whereas in the model-based case the policy should be updated and the animal should avoid that state. Both electrophysiological and fMRI measurements in such a setup could show areas that seem to encode the devaluated reward [Bornstein & Daw 2011, Dolan 2007, Gottfried *et al.* 2003, Matsumoto & Tanaka 2004, Valentin *et al.* 2007].

However early studies showed already a strong influence of the experimental setup on the success of reinforcer devaluation [Adams 1982, Dickinson 1985]. Specifically, animals seem to act model-based after short training schedules, but model-free after extended practice. Those findings fit well with much older research noting that new tasks require high effort by the subjects, but after extended practice are performed with ease [Bryan & Harter 1897, James 1890]. This lead to the proposal that both systems are used by the brain (for a review see [Sloman 1996]). Later studies tried to address the physiological substrates of the two systems and how the selection or integration between them works. Killcross and Coutureau for example showed that a lesion in the dorsal part of medial prefrontal cortex (PFC) made rats insensitive to reinforcer devaluation independent of training time. Lesioning the ventral part of PFC in contrast resulted in persistence of model-based behaviour even after extended training [Killcross & Coutureau 2003]. A similar study found a knock-out of the model-free mechanisms after lesions in dorsolateral striatum [Yin *et al.* 2004]. Despite showing an anatomical separation between the two behavioural controllers, knowledge about the implementation of a selection mechanism is still missing. A recent paper by Daw and colleagues found evidence for a shared representation at least of the prediction errors [Daw *et al.* 2011], pointing towards the possibility of an integration before the decision process.

Beyond that, studies tried to shed light on the specific instantiation of model-free learning. By now those results are still mixed, finding correlates with SARSA variables in one case [Morris *et al.* 2006] or Q-learning ones in another [Roesch *et al.* 2007]. There is also a potential separation of an actor and a critic seen in brain activity [O'Doherty *et al.* 2004].

Theoretical work has also tried to address the question why the brain uses multiple systems and how it could choose which one to use in a given task [Daw *et al.* 2005, Dayan & Daw 2008, Keramati *et al.* 2011, Lengyel & Dayan 2008, Shah & Barto 2009, Summerfield *et al.* 2011]. The main idea behind most of the theories is to optimize not only the collected reward but also an additional value. Daw and co-workers proposed for example that the winning controller will be the one that has smaller uncertainty about the predicted values [Daw *et al.* 2005]. In the model-based controller, this value uncertainty results from uncertainty in the environmental structure and potentially additional noise from using approximation methods to compute the value function in reasonable time. The model-free controller's value function is uncertain due to the local averaging of past experiences. In a simulation of the devaluation experiments they showed that at the start the model-based controller has less uncertainty in his value function, but after extensive training the model-free system can become superior. Experiments using fMRI on humans doing a specifically designed task did find evidence for this model by showing activity correlated with one or the other controller in multiple areas [Beierholm *et al.* 2011]

In an extension of this work Keramati and colleagues proposed a more implicit formulation of these uncertainty values [Keramati *et al.* 2011]. They computed an average reward rate per time and a function they call the "value of perfect information (VPI)", which compares the probability functions for rewards of different actions and the more similar they are, the more it helps to know the true value. As long as the VPI is higher than the reward rate, the agent will use the model-based controller. If its increased processing time results in missing many action opportunities with potential reward and can not be compensated by an increase in the number of actions that are rewarded, the model-free system will take control over behaviour. This is shown to happen when training time increases and therefore also qualitatively reproduces the results from devaluation experiments.

There is even recent work proposing a third, episodic, controller, which uses stored state-action sequences that once led to a reward and which could be beneficial in the very early phase of training [Lengyel & Dayan 2008].

Generally it can be said, that there is not one single system in the brain that controls behaviour. The theory proposed in this thesis is addressing learning in the model-free system and is not excluding other approaches targeting the model-based controller. The idea of multiple controllers could also be a possibility to explain both fast learning [Triesch *et al.* 2002] and strong stability [Michel & Jacobs 2007] for artificial setups in cue integration.

3 Bayesian Cue Integration as a Developmental Outcome of Reward Mediated Learning

3.1 Introduction

In this chapter I will present a model for the development of cue integration that will lead to performances that match those predicted by an optimal Bayesian observer. The model tries to provide an explanation for the finding that children, unlike adults, seem to not be able to use multiple information sources in a way that leads to near-optimal outcomes. The work in this Chapter tries to address the *algorithmic level* of theoretical research on the topic, which means it will show that reward-mediated learning can be one mechanism that could explain the psychophysical results. We even go beyond basic cue integration by additionally modeling the development of causal inference, another interesting topic related to perceptual inference. The concrete implementation that is used is more abstract and not proposed to be the implementation used in the brain – this question will be discussed in Chapter 5.

3.2 Methods

We use a multimodal localization task similar to the one used by Neil and colleagues [Neil *et al.* 2006] and Körding et al [Körding *et al.* 2007] (see Fig. 3.1 for a schematic depiction). The learner obtains noisy visual and auditory signals and carries out horizontal orienting movements, obtaining a varying amount of reward depending on the accuracy of the movement (see 3.2.1). We interpret the reward as an intrinsic signal for bringing a relevant stimulus into the center of attention. Orienting movements were shown to be adaptive and sensitive to reward [Hikosaka *et al.* 2006, Platt & Glimcher 1999, Takikawa *et al.* 2002, Schultz 2000].

The agent learns to solve this task based only on its sensory inputs, orienting actions, and observed rewards. To this end, it learns to predict how much reward to expect when performing each action in a given situation. The learner represents its reward estimates for particular state and action pairs as Q-values [Sutton & Barto 1998] (see Section 2.3.1). Support for the representation of such variables in the human and monkey brain comes from several studies [Morris *et al.* 2006, Samejima *et al.* 2005]. In our case this Q-function is approximated by a three-layer artificial neural network (ANN) (see 3.2.2), based on the noisy sensory position estimates. Using these reward expectations, the agent will probabilistically pick an action according to a softmax function (eq. (3.1)), which also has been shown to match human action selection for some tasks [Daw *et al.* 2006, Rangel & Hare 2010]. The reward prediction of the winning action will be adapted depending on the difference between predicted and obtained reward by changing all synaptic weights via a gradient descent learning algorithm (see Subsection 3.2.2).

3.2.1 Task Setup: Simulations

In our task each trial consists of the presentation of two stimuli in the visual and auditory modalities. These stimuli either originate from a single common source (Fig. 3.1 left) for the auditory and visual cue or from two separate sources/objects (Fig. 3.1 right).

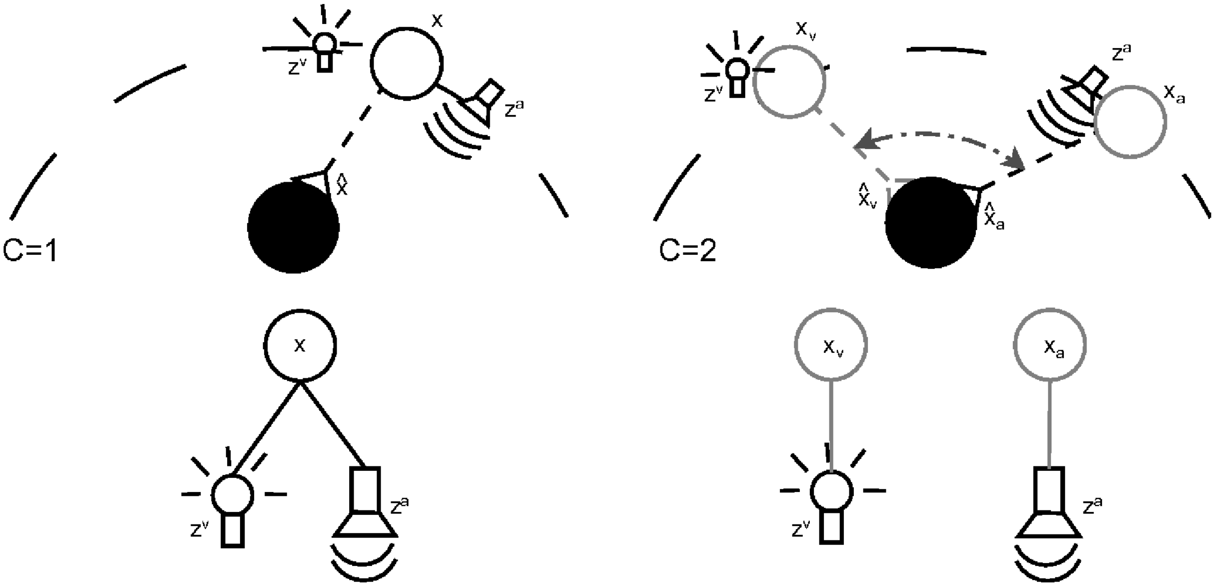

Figure 3.1. Scene layout of orienting tasks and generative models. Top: Sketch of the orienting task used in this study. The learner receives an auditory (z^a) and a visual (z^v) signal, which are probabilistically related to the true position x. The task is to orient towards this true position. Bottom: The generative models for the task. The left side shows the case where the visual and auditory signals have a common cause ($C = 1$), On the right the signals originate from different locations ($C = 2$).

A position x along the spatial dimension is chosen from a uniform distribution for each object in the scene (but results if, e.g., the prior for visual position is Gaussian around the central region are not different – Table 3.1 1C). In the two objects case we call their positions x_a (for the object that emits only an auditory signal) and x_v (for the object that emits only a visual signal). Space is discretized to $x_{\max} = 30$ positions for ease of computation. The received sensory signals are noisy versions of the true source locations. We use additive noise with normal distributions with zero mean and variances σ_a^2 and σ_v^2. Note that the reinforcement learning (RL) model is also able to deal with noise from different distributions since we do not implement the learner based on a fixed distribution. See Table 3.1 2B for a setup with auditory and visual noise drawn from a logistic distribution with median 0. This noise is thought as being of sensory and/or environmental origin, e.g. background noise, neuronal firing stochasticities and tuning densities. Usually the variance of the auditory estimate is set larger than the visual one, in accordance with psychophysical observations for spatial tasks [Thomas 1941]. We call this noisy signal position z^a and z^v respectively. If the noise makes a signal fall outside of the spatial range, the stimulus is treated as not present, thus resulting in a unisensory training trial. An important implication of this setting is that the structure of observations is the same for both possible underlying generative models.

We use two slightly different versions of this task. In the single output task the learner has to orient towards a single location. That means in the case of two objects the reward only depends on the distance to the object closest to the estimated position. In the two outputs task it is required to orient towards both the visual and the auditory positions of their respective cause. In case of a common cause this should result in both estimates being equal. There are separate rewards for the visual and auditory action. The inputs were the same for both experiments.

3.2.2 Reinforcement Learning Model

An approximation of the function relating state-action pairs to predicted reward is learned. A three-layered ANN (see Fig. 3.2) is set up with an input unit for each position in every modality (here 60 input neurons). It should be mentioned that the yet unsolved problem of limited scalability of RL approaches for very large numbers of inputs, does also apply to our model. The input neurons i are all-to-all connected with weights $v_{i,j}$ to neurons j in the hidden layer (here $j = 0...29$). Stimulus locations z^a and z^v are represented by the population activity of these input neurons (see e.g. [Fuzessery *et al.* 1985, Lee *et al.* 1988] for biological examples of population codes) in each modality separately (the first half of the neurons coding for the auditory input, the second for the visual one).

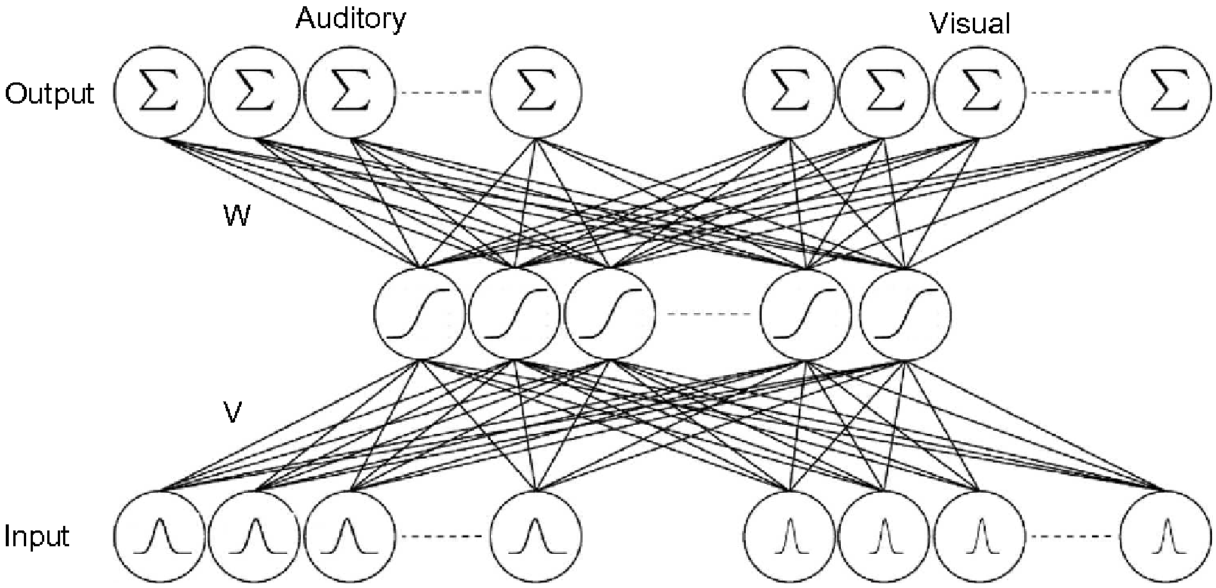

Figure 3.2. Sketch of the neural network used for approximation of the Q-value function. Setup for the two–step orienting task, the setup for the simple orienting task differs only in that the network has only half as many output neurons, since only a single action is required.

Each input neuron has a Gaussian receptive field, centered on position z^a_i or respectively $z^v_{i-x_{\max}}$ for $i \geq x_{\max}$. The variance of these Gaussians is of the order of the noise of the input stimuli. Overlapping receptive fields of the input neurons simply help the network to discover a spatial relationship between the possible input positions. We also tested the framework with simple binary input units and found no difference in the final results apart from an increase in learning time (see Table 3.1 1B).

A sigmoidal transfer function on the sum of the weighted inputs g_i produces the activation y_j of the hidden neurons. These are again fully connected to the output neurons x' with weights $w_{j,x'}$. For every action x' there is one output unit, with its activation $o_{x'}$, given by the weighted sum of the hidden layer activity, representing an approximation of the appropriate Q-value. All weights are initially drawn from a uniform distribution, the v's between -0.1 and 0.1, the w's between -1 and 1.

Based on the network's outputs the learner chooses one of the available actions. This is done with the softmax function:

$$P(\hat{x} = x'|s) = \frac{e^{Q_{x',s}/\tau}}{\sum_{\overline{a}} e^{Q_{\overline{a},s}/\tau}}. \tag{3.1}$$

This probabilistic action selection rule chooses an action x' with a probability proportional to the relative predicted reward $Q_{x',s}$ for that action, given state s. We start with a high temperature parameter $\tau = \tau_0$, so that the learner chooses his actions only weakly influenced by the initial reward expectations. τ then decreases exponentially with learning time (with $\tau(t) = \tau_0^{\frac{\nu_\tau - t}{\nu_\tau}}$), passing 1 after a given number of steps ν_τ. At smaller values of τ the selection favors more and more the action with highest expected reward, thus exploiting the environment.

After performing the selected action $\hat{x}$, the learner receives the true reward $r(\hat{x})$. We use a reward function that is maximal if $\hat{x}$ equals the true object position x, decaying quadratically with increasing distance within a surrounding area (with radius ρ) and zero otherwise.
For single output:

$$r(\hat{x}|x_a, x_v) = \max(0, (\rho - \min(|x_a - \hat{x}|, |x_v - \hat{x}|)))^2 \tag{3.2}$$

For two outputs:

$$\begin{aligned} r_a(\hat{x}_a|x_a, x_v) &= \max(0, (\rho - |x_a - \hat{x}_a|))^2 \\ r_v(\hat{x}_v|x_a, x_v) &= \max(0, (\rho - |x_v - \hat{x}_v|))^2 \end{aligned} \tag{3.3}$$

If only one object is present, the two position are equal, i.e. $x_a = x_v$. In the experiments shown above we used $\rho = 4$. Changes of ρ other than setting it to zero (only rewarding correct actions) only have an impact on the learning time. We also tested the model with an asymmetric reward function, where a correct visual action would only provide half the reward of a correct auditory action (results see Table 3.1 2C).

Based on the true reward, the Q-value for the particular state-action pair will be updated proportional to the difference between prediction $Q_{\hat{x},s}$ and $r(\hat{x})$. This difference can be seen as a temporal difference (TD) error for a single time step. TD learning in general uses discounted future rewards for computing the prediction error: The Q-value function will not only represent the expected reward of a single state–action pair, but also include possible future rewards that are expected from the new state. In the present work the learner has to only perform a single action per trial and receives only immediate reward.

To minimize the TD error we use gradient descent to change the weights of the neural network [Rumelhart *et al.* 1986b, Rumelhart *et al.* 1986a] with:

$$\Delta w_{j,x'} = \begin{cases} -\varepsilon(r_{\hat{x}} - o_{\hat{x}})(-y_j), & \text{if } x' = \hat{x} \\ 0, & \text{else} \end{cases} \tag{3.4}$$

$$\Delta v_{i,j} = -\varepsilon(r_{\hat{x}} - o_{\hat{x}})(-w_{j,\hat{x}})y_j(1 - y_j)g_i. \tag{3.5}$$

ε is an exponentially decreasing learning rate: $\varepsilon(t) = 10^{\log(\varepsilon_0) - \frac{t}{\nu_\varepsilon}}$, with $\varepsilon_0 = 0.05$ and $\nu_\varepsilon = 100,000$. The results did not change much when using an alternative function for the learning rate, $\varepsilon(t) = \frac{\varepsilon_0}{\text{ceil}(\frac{t}{\nu_\varepsilon})}$, with $\nu_\varepsilon = 10,000$.

3.2.3 Bayesian Observer Models

We compare the performance of our model with that of four different Bayesian observers, inferring the position of the object given the input and the generating model C (Fig. 3.1 bottom). With Bayes' theorem and the assumption that the noise of different modalities is independent we can write the posterior probability as:

$$p(x|z^a, z^v) = \frac{p(z^a, z^v|x)p(x)}{p(z^a, z^v)} = \frac{p(z^a|x)p(z^v|x)p(x)}{p(z^a, z^v)}. \tag{3.6}$$

where the last equality is only valid if the two cues are conditionally independent given their cause. The likelihoods $p(z^a|x)$ and $p(z^v|x)$ include all information available from the input. The reliability of a cue is inversely proportional to the standard deviation of this distribution. In the experiments reported here the prior $p(x)$ is always uniform. Other priors were used in simulations, and the RL algorithm was able to adjust to these and still perform close to the Bayesian predictions (Table 3.1 1C for an example with $p(x_a) = \mathcal{N}(x_a, 15, 7.5)$). Since we are interested in the performance of the model in terms of reward, actions are not chosen only based on the posterior probabilities, but on the utility function $U(\hat{x}|z^a, z^v)$, which additionally takes into account the expected reward $r(\hat{x}_{[a,v]}|x_a, x_v)$ (we write $[a, v]$ to cover both the one and two output case) for a given action (see below). The use of different utility functions can accommodate different tasks in a very direct way and makes the behavioral goal explicit.

The Bayesian observers used here differ in the way they handle the two different possible generative models (one vs. two causes; Fig. 3.1 bottom). model averaging (MA) uses a utility function that is a weighted average of the inference results of each model. The weights are determined by the probability for one versus two objects $p(C|z^a, z^v)$. This probability can again be computed from known distributions using Bayes formula (2.1).

$$U(\hat{x}_{[a,v]}|z^a, z^v) = \begin{array}{c} p(C = 1|z^a, z^v) \int r(\hat{x}_{[a,v]}|x)p(x|z^a, z^v)dx \\ +p(C = 2|z^a, z^v) \iint r(\hat{x}_{[a,v]}|x_a, x_v)p(x_a|z^a)p(x_v|z^v)dx_a dx_v \end{array} \tag{3.7}$$

model selection (MS) in contrast uses only the utility function of the most probable model.

$$U(\hat{x}_{[a,v]}|z^a, z^v) = \left\{ \begin{array}{ll} \int r(\hat{x}_{[a,v]}|x_a, x_v)p(x|z^a, z^v)\, dx, & \text{if} \quad p(C = 1|z^a, z^v) > 0.5 \\ \iint r(\hat{x}_{[a,v]}, x_a, x_v)p(x_a|z^a)p(x_v|z^v)\, dx_a dx_v, & \text{else} \end{array} \right. \tag{3.8}$$

We use a uniform prior over the number of objects in the scene ($P(C = 1) = P(C = 2) = 0.5$). Results of additional simulations not shown here lead to similar results for asymmetric distributions.

We also consider two observers that only do inference on one model, ignoring the second one – one model always integrating the cues (AI) and the other model always treating the cues as independent – never integrating (NI). The utility functions of all observer models are computed by numerical integration. For a given input we choose the action with maximum utility.

Another possible observer model would compute the same probability distributions as in MS and MA, but then select stochastically from them instead of choosing the maximum. Such a behavior is often called probability matching (PM). In our case it could be used in two ways: A recent paper proposed PM at the level of causal inference [Wozny *et al.* 2010], an action will be chosen according to one of the generating models with the probability for that model to be the underlying cause ($P(C)$). Because this is an intermediate between MS and MA we only consider it when computing the R^2, where we distinguish between those. The second possibility would be to use PM for the action selection step, which was found in various studies to be a strategy employed by human observers in certain tasks [Grant *et al.* 1951, Rubinstein 1959]. This is actually implicitly assumed in our model by using the softmax function to pick the action, thereby we do not include this option in our analysis.

3.3 Results

In the following we will test our model on the cue integration and causal inference tasks and compare it to human behavior and four different Bayesian models. Most of the simulation results in this Chapter were published in [Weisswange *et al.* 2011], preliminary reports can also be found in [Weisswange *et al.* 2009a, Weisswange *et al.* 2010]. A first work on this topic but using tabular RL can be found in [Weisswange *et al.* 2009b], but is not explicitly referred to in the results section.

3.3.1 Cue Integration

We start with a simple cue integration paradigm, where noisy auditory and visual signals from a common source have to be combined. If the noise of the two cues is independent, the variance of the error produced by optimally integrating the two stimuli is always smaller or equal to the error variance resulting from using either cue alone. Figure 3.3 shows the distribution of errors the RL based model produces after training. This result matches well with the predictions of the optimal Bayesian model for this situation.

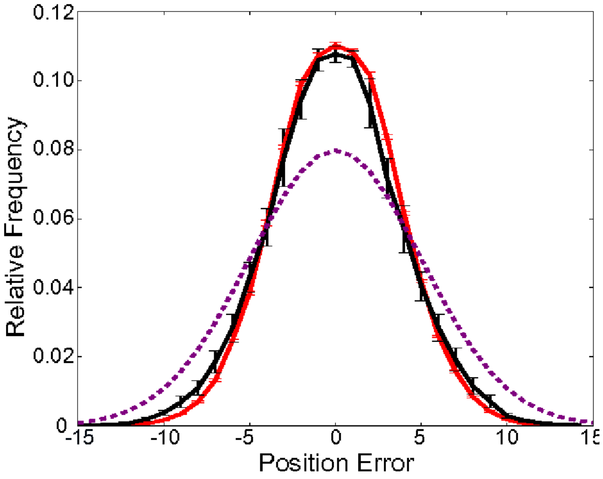

Figure 3.3. Distribution of position estimation errors. The distribution of errors over 100,000 orienting actions carried out by the RL model after 10 runs with each 100,000 training steps (black), compared with Bayesian optimal integration (red) and the best single cue predictions (dashed) for a single audio-visual object. Error bars show standard deviation of 10 runs. ($\sigma_a^2 = 5$, $\sigma_v^2 = 5$)

To compare the model with human behavior, we test the fully trained model on a two-alternative forced choice (2-AFC) task. This task allows us to test the behavior of the learner for changes in relative reliabilities between the cues. The setup is similar to the one used by Ernst and Banks [Ernst & Banks 2002], where human subjects were asked to perform a 2-AFC visuo-haptic size discrimination task. Ernst and Banks could show that in this task the point of subjective equality (PSE) of adults is well predicted by Bayesian cue integration and shifts when additional visual noise is introduced.

The first input to the agent is the size of a standard object with constant position, the second is the size of a probe which varies and is to be estimated as 'left' or 'right' of the standard (respectively 'taller' or 'smaller' in [Ernst & Banks 2002]). Both stimuli are bimodal, but for the probe the cues are always consistent, whereas for the standard they are set to be in conflict with each other. Figure 3.4A shows the proportion of 'on the right' estimates for all possible positions of the probe based on the decisions taken by the reinforcement learner after training as psychometric curves. Each curve represents training and testing with a different visual noise variance. We can compare it with the data from of Ernst and Banks [Ernst & Banks 2002] (Figure 3B) which is reproduced in our Figure 3.4B. It shows the equivalent data for the average of four human subjects. Both plots show a similar pattern in that the psychometric curves get steeper and the PSE moves more towards the visual stimulus position for decreasing visual noise levels.

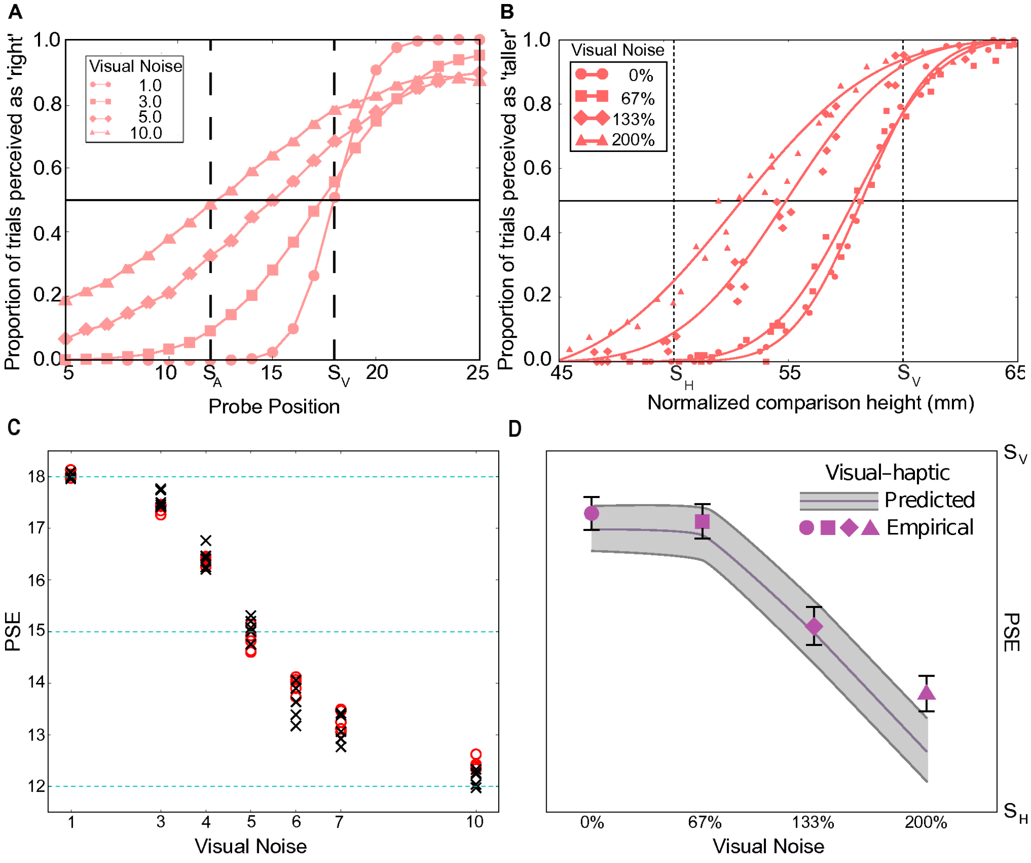

Figure 3.4. Psychometric curves and PSEs of the model for the 2-AFC task in comparison to human psychophysics. A: Psychometric curves for the proportion of 'on the right'-actions in an audio-visual 2-AFC position estimation task. The input positions of the standard were mismatched, with the auditory signal positioned at 12 and the visual signal at 18. Probe inputs were matched and tested 1,000 times at each position. The point at which the curves cross the black horizontal line is the PSE. The curves differ in the variance of the visual noise (see legend), auditory noise was kept constant with $\sigma_a^2 = 5$. B: Plot using data from a psychophysical experiment by Ernst and Banks [Ernst & Banks 2002]. They used a visuo-haptic 2-AFC size discrimination task and count the proportion of 'taller'-actions. The standard inputs were mismatched (haptic at 50, visual at 60), probe inputs were matched and varied between 45 and 65. The visual reliability was varied by adding external noise to the display. C: Plot of the PSEs of the RL model (black crosses) and Bayesian integration (red circles) for each of 5 repetitions of training and 2-AFC testing with different values for the variance of the visual noise σ_v^2. For all trials $\sigma_a^2 = 5$. In the test trials the visual and auditory signal of the standard were positioned like in A (dashed blue lines). The variance in the Bayesian data points are results of limited sample size. D: Plot adapted from psychophysical experiments by Ernst and Banks [Ernst & Banks 2002] showing the change of the PSEs of human observers for different levels of visual noise in a visuo-haptic discrimination task. The gray area shows the predictions from a Bayesian model.

Figure 3.4C compares the PSEs of the RL model (crosses) with that of the optimal Bayesian observer (circles) for different visual reliabilities. It can be seen that they match quite well, as was true for the human subjects in [Ernst & Banks 2002] (Fig 3.4D). Note that there is variability in both the PSEs of the learner and the Bayesian observer due to the limited number of test stimuli.

3.3.2 Causal Inference

In the following tasks we will add a second layer of complexity by randomly presenting trials that were generated by different scene layouts, i.e. under either the common or the separate cue condition. We will compare our learned model with four Bayesian observers. One observer always integrates the information from the two stimuli (AI). A second always acts as if both stimuli originate from different objects and discards information from the less reliable modality (NI). A third, more advanced, observer computes the probability of one vs. two objects in each trial and uses the optimal action for the more probable model (MS). The fully optimal fourth observer though makes use of all information available by selecting an action under the weighted evidence for each generative model, with the weights proportional to the respective probabilities (MA). All Bayesian observers, contrary to the RL model, have explicit knowledge of position priors, sensory noise distributions and the reward rule. The mathematical formulation of these decision rules as well as the reward expectations of the observer models can be found in Section 3.2.

To show the learning process of the RL learner, we can look at the development of the potential reward received with a greedy policy (always selecting the action which predicts the highest reward). In Figure 3.5 one can see that the average reward earned by the learner increases until it reaches a level similar to what the MA and MS models show (see also Table 3.1 1A). Comparing it with the simpler instances of Bayesian observers, the learner is clearly better than AI and NI, that is, it implicitly incorporates the existence of two different conditions. But it is hard to tell apart the Bayesian MA and MS observers. Both are similar to the agent's performance in this task.

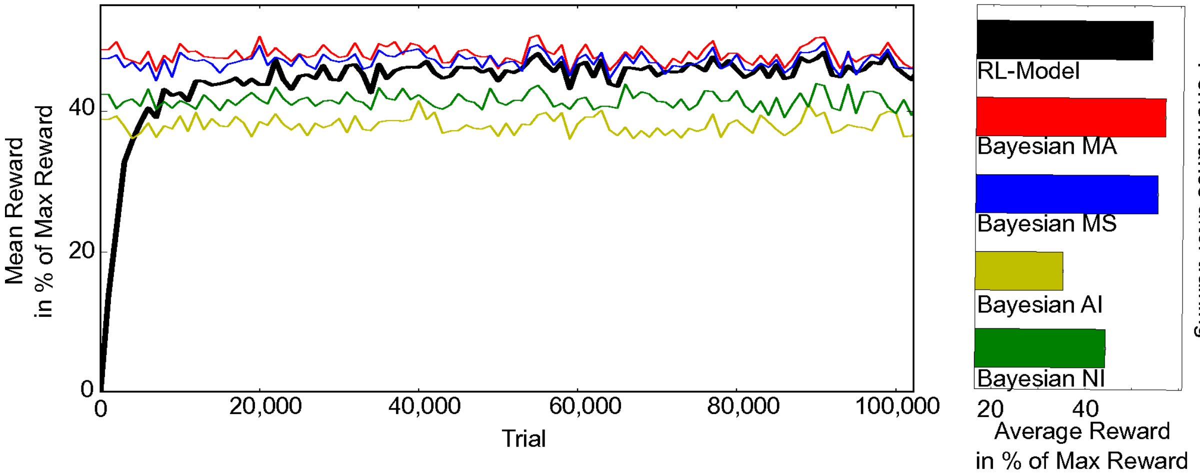

Figure 3.5. Performance of the RL model and Bayesian observers for a single output. Reward obtained by the learner when choosing the action with highest predicted reward (black) compared to the different Bayesian observers. Signals can originate from one or two objects. Left: Change of performance during learning. Each data point is the sliding average of 1000 trials. Right: Bar plot of the mean reward over 100,000 trials after learning. Standard error of the means is smaller than 0.5% for all bars. ($\sigma_a^2 = 3$, $\sigma_v^2 = 2$)

A different way of assessing the behavior of the RL agent is to directly consider the expected total discounted reward obtainable for a particular state-action pair, the Q-value. Figure 3.6 shows subsections of two learned Q-value function approximations for all inputs, a given action and two different reliability ratios. The highest reward is expected if both input signals are close together, resulting in a high probability for a single cause, and close to the target of the given action, resulting in a high probability for the action being correct (Fig. 3.6 center of both plots). Importantly, if the target of the given action can not possibly be a result of a weighted average of the input positions – because the cues favor both a higher or both a lower position, this action predicts little reward (asterisks in Fig. 3.6). For this reason the plots show an asymmetric reward landscape. The slant of the area of highest reward (dark red) depends on the relative reliability of the two cues, as can be seen when comparing A and B in Fig. 3.6. The left plot is a result of inputs with a higher reliability in the visual modality, therefore the area of highest reward lies more along the visual axis, whereas in the right plot with equal reliabilities for both cues it lies along the diagonal exactly between the auditory and visual axis. The width of this area, as well as the maximum predicted reward, is determined by the absolute values of the reliabilities (narrower and darker red in the left plot due to higher visual reliability). A smaller reward can be expected if the cues are far apart – resulting in a high probability for two causes, but one of them is close to the action target – resulting in a high probability for the action to be correct for one of the objects (Middle of each of the four figure boundaries in Fig. 3.6 – the "arms" of the cross). The height of these expectations depends again on the reliability of the relevant cue.

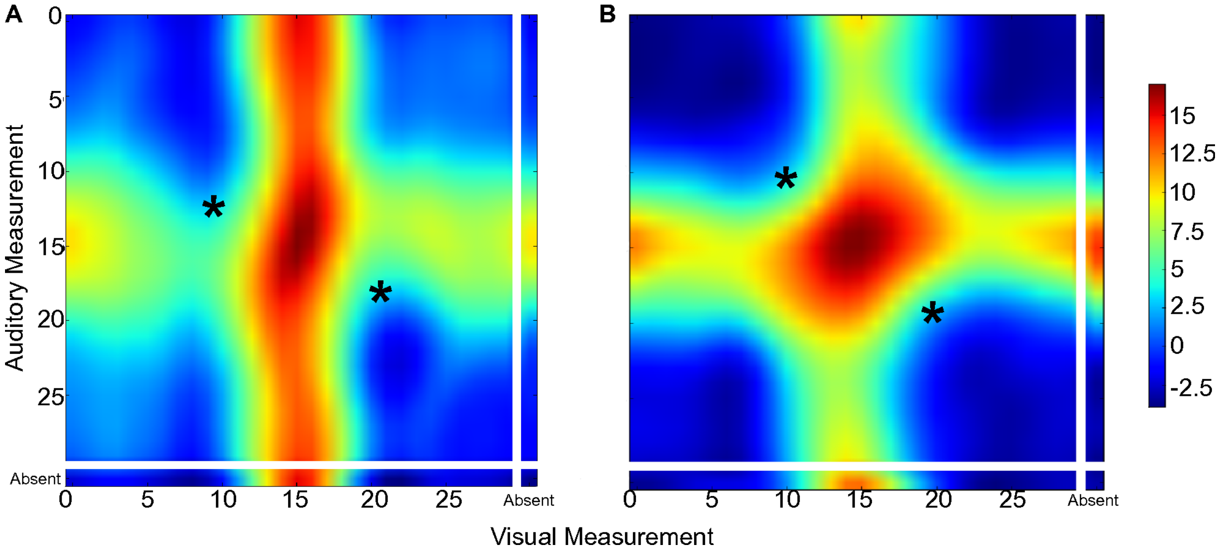

Figure 3.6. Exemplary subsections of two learned Q-value functions. Expected reward for visual signals (x-axis) and auditory signals (y-axis) for the action of orienting towards the center. Red colors represent high, blue low predicted rewards. Left: Visual cue is more reliable ($\sigma_a^2 = 3$, $\sigma_v^2 = 2$); Right: Both cues have the same variance ($\sigma_a^2 = \sigma_v^2 = 3$). For a detailed explanation see main text.

In the experimental setup from [Körding *et al.* 2007] participants were asked to report in each trial both the visual as well as the auditory location of a stimulus. To mimic this condition, we change our task accordingly and add a second output population to the neural network (see Fig. 3.2). Each population now represents the actions for one modality (representing the arrays of buttons for the participants in [Körding *et al.* 2007]). The rewards and the prediction errors are computed separately (according to (3.3)). Table 3.1 2A shows the performance after learning as the sum of both rewards. The effects are similar to the previous orienting task, in that we see a performance that is similar to the predictions of

MA and MS.

Table 3.1. Model performance for different set-ups

Setup	RL-Model	Bayesian MA	Bayesian MS	Bayesian AI	Bayesian NI
1A	46.62%	47.9%	47.04%	37.87%	41.57%
1B	46.32%	47.9%	47.04%	37.87%	41.57%
1C	45.86%	47.06%	46.18%	37.28%	40.88%
2A	50.51%	51.81%	51.08%	41.63%	47.71%
2B	36.26%	37.37%	36.53%	33.87%	34.28%
2C	48.74%	50.52%	49.70%	40.32%	45.67%

Average fraction of maximum reward received in 100,000 steps after learning ($\sigma_a^2 = 3$, $\sigma_v^2 = 2$) for different variations of the task and the model. Results of the different Bayesian observers for comparison. All setups use $p(C = 1) = 0.5$. The setups with references starting with 1 are using a single output, those with a 2 use one auditory and one visual output. 1A and 2A are the standart setup with the two types of outputs. 1B uses inputs encoded as a binary instead of a population code vector, in 1C the visual objects have a Gaussian prior on their position which is centered at 15. In 2B auditory and visual noise are drawn from a logistic (instead of Gaussian) distribution with median 0. Finally 2C uses an asymmetric reward function, where the reward for a correct visual response is only half of that of the auditory one.

Despite changing the task it is still difficult to distinguish these two Bayesian observers (MA and MS) from each other by comparing the collected reward. The main reason for this similarity can be seen in Fig. 3.7, which shows the policies of the two observers. They only differ for a very narrow range of stimuli, which in addition are also very rare during normal training/test runs.

A better discriminator should be the variance explained by each observer in relation to the total variance of the orienting error of our model (generalized coefficient of determination R^2 [Rao 1973]). The differences between MA and MS over all inputs are nevertheless still small (Fig. 3.8).

Fortunately, since we have the full observer models, we can find the inputs for which the optimal actions differ between MA and MS, and then test the RL model only on those (Fig. 3.7B). The R^2 values for these inputs are shown in Fig. 3.9. We also include an observer which does PM for model selection, proposed to be the strategy used by many human subjects in a recent experiment [Wozny *et al.* 2010]. It can be seen that the Bayesian observer with MA explains the error variances best for both visual and auditory output (gray and black bars). The values for the MS and NI observer are the same, because the selected inputs represent those in which MS decides to act according to the generating model with independent objects.

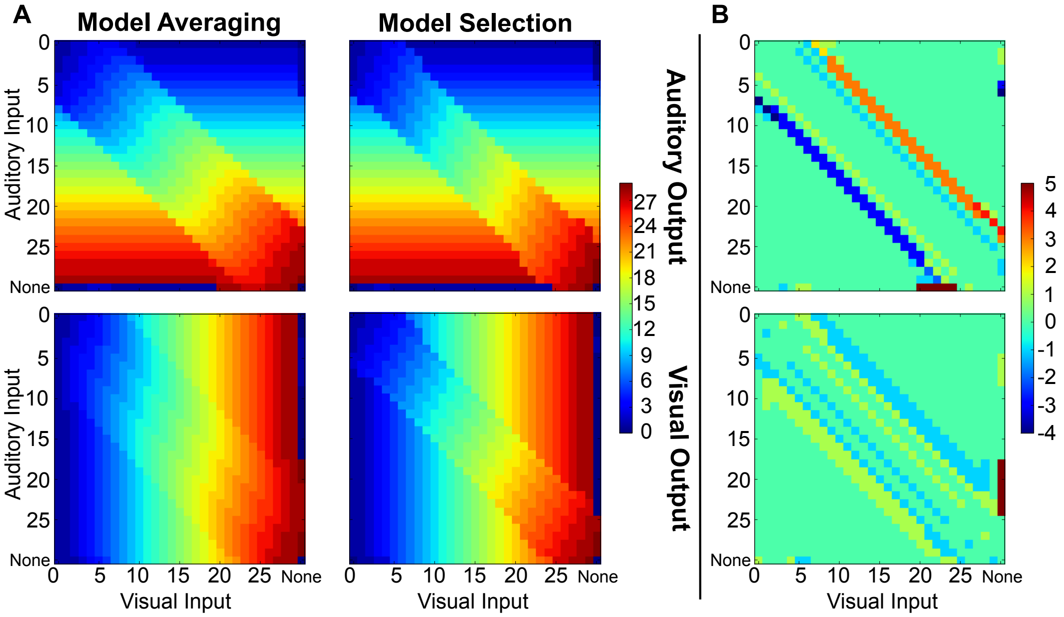

Figure 3.7. Policies of the MA and MS observers for two outputs. A: Plots show the MAP estimates of position (color coded) for the auditory (top row) and visual (bottom row) signal from the two Bayesian observers MA and MS. **B**: The difference between the MA and MS policies. The output predictions only differ in the small border region where the probabilities for one vs. two causes are very similar. ($\sigma_a^2 = 3$, $\sigma_v^2 = 2$)

3.3.3 Complex Uncertainty Structures

After showing the general ability of the system to do cue integration and causal inference, we are also interested in its robustness against changes in the distributions used to generate the inputs. The system can for example accommodate different prior distributions of the scene variables relevant for obtaining rewards (see Table 3.1 1C for an example with a Gaussian prior for the visual stimulus) Beyond that it is interesting to test whether it can also handle different likelihood landscapes. In many real-life situations, the uncertainty of a cue is influenced by a variety of factors. The following three experiments introduce behaviorally plausible variations in uncertainty structure and investigate how the RL agent can adjust to these.

Spatial Variation in Uncertainty Structure

Visual estimates of spatial location should be more accurate in the fovea than in the periphery of the visual field, given the human acuity falloff (e.g. [Hairston *et al.* 2003], see also [Knill 2005] for an example in slant angle space). Figure 3.10 shows the reward predictions for a set-up that mimics this observation in the task that requires a single action. The variance of the visual noise was low for stimuli in the center and increased with eccentricity, whereas auditory reliability stayed constant (Figure 3.10 shows results with linear increase of the variance, but similar results are reached with other functions, e.g. logarithmic decay). Training on this adapted task resulted in reward predictions dominated by the visual estimate for actions towards the center (Fig. 3.10 right) and dominated by audition for the outer periphery (Fig. 3.10 left).

In between these two extremes, integration of both cues predicted the highest reward. This can also

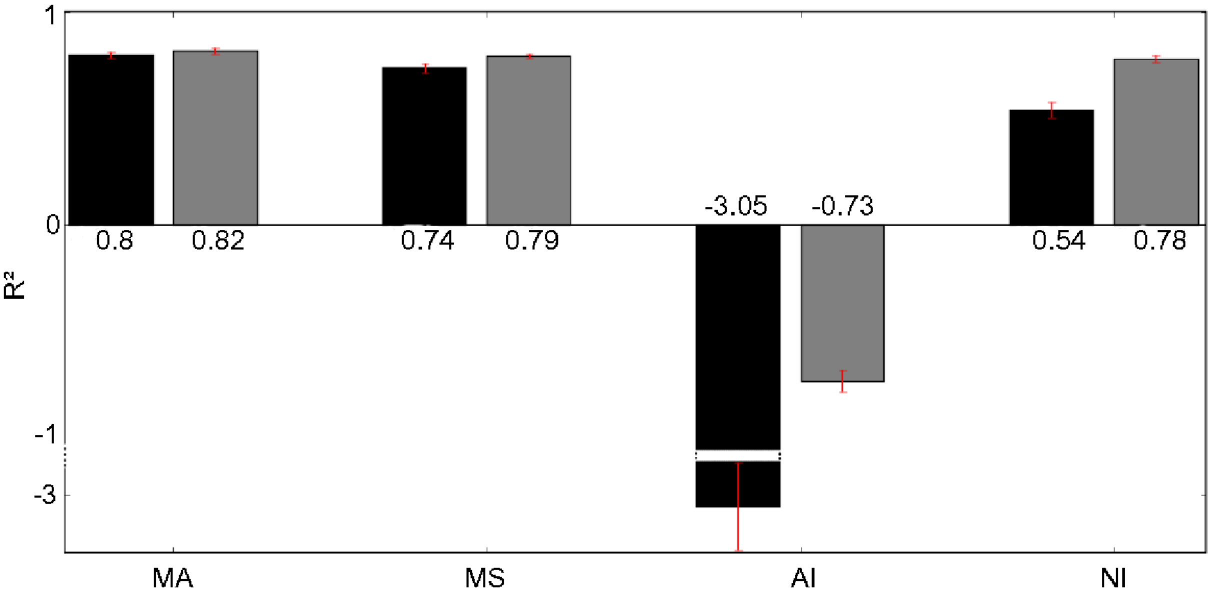

Figure 3.8. R^2 values of different observers for the responses of the RL model for all inputs. The black and gray bars show the results for the auditory and the visual output from the complete space for 50,000 trials. Mean over 10 training sessions with $\sigma_a^2 = 3$, $\sigma_v^2 = 2$, error bars show standard deviation.

be seen in the distribution of input weights to the hidden layer (Fig. 3.11). The weights from the auditory part of the input layer have similar shapes and width across all positions, whereas the visual weights get narrower towards the central positions. This shows that reward mediated learning results in behavior that varies with context within a single task, which is in accordance with predictions from a Bayesian model that explicitly takes into account context when computing the data likelihood.

Temporal Variation in Uncertainty Structure

In addition to a change in noise variance across space as discussed above, in a natural environment the variance also changes over time. As an example one may consider the change in the optimal weighting of visual compared to auditory cues when stepping out of a dim hallway into a well lit room. Due to higher contrasts and thus smaller uncertainty, visual localization will gain confidence in the latter condition. To simulate such dynamics, we change the reliability of the visual cue at certain time points during training (Fig. 3.12). The network quickly adjusts to a change in visual reliability. The performance after a change point (vertical lines in Fig. 3.12) quickly becomes similar to the optimal predictions by the Bayesian observers. This is mostly due to the generalization abilities of the function approximation. A learner using a table with entries for every state-action-reward mapping [Weisswange *et al.* 2009b] has to effectively relearn its policy with every change in conditions. It should be noted that the Bayesian observers were explicitly given the new uncertainty distributions after every change, and because of that could react instantaneously.

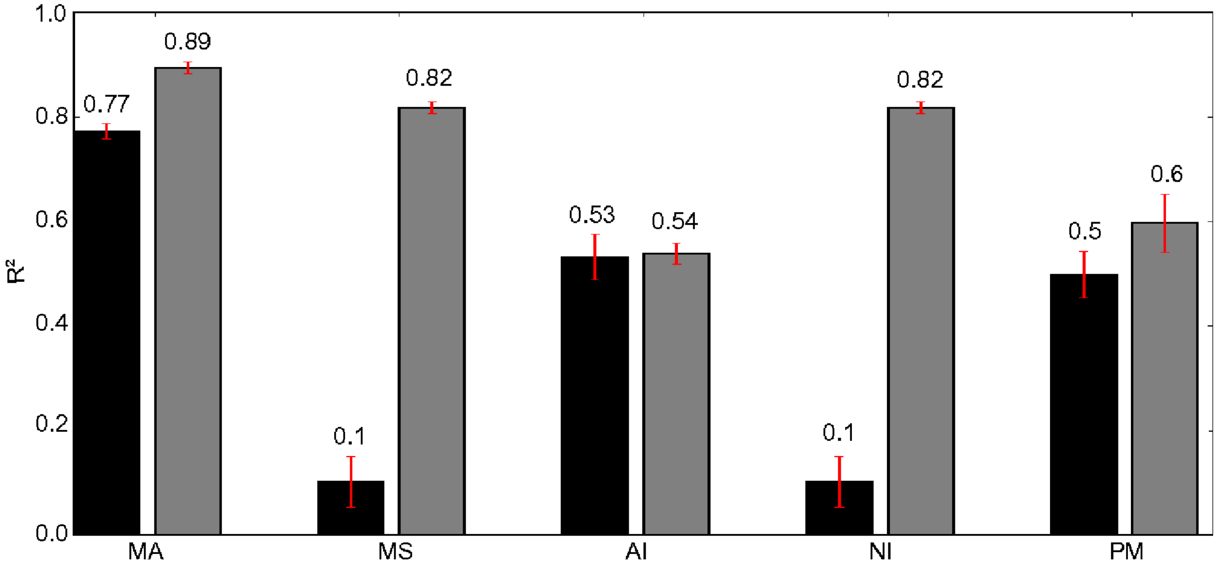

Figure 3.9. R^2 values of different observers for the model with selected inputs. The black and gray bars show the results for the auditory and the visual output for 50, 000 trials with inputs that differ in the predicted action between MA and MS. Mean over 10 training sessions with $\sigma_a^2 = 3$, $\sigma_v^2 = 2$, error bars show standard deviation.

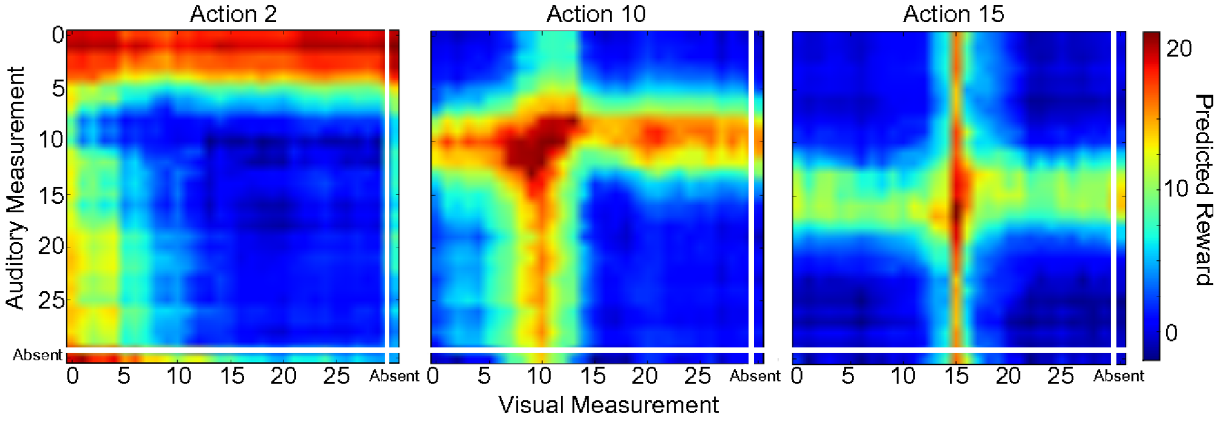

Figure 3.10. Exemplary subsections of the learned Q-value function for the foveation setup. Axes are the same as Fig. 3.6, but for three different actions with constant auditory reliability ($\sigma_a^2 = 2$) and space varying visual reliability. L-R: The actions of moving towards a peripheral ($\sigma_v^2 = 3.25$), intermediate ($\sigma_v^2 = 2$) and central position ($\sigma_v^2 = 0.25$) are shown.

Shift in Uncertainty Structure

We can also adapt our settings to simulate the conditions used in the experiment by Wallace and Stein [Wallace & Stein 2007] to introduce mismatches in the spatial alignment of stimuli from a common object. We ask whether reinforcement mediated learning could also produce results similar to the aberrant spatial integration found in their study. Therefore we bias the auditory signal by setting the mean of its noise

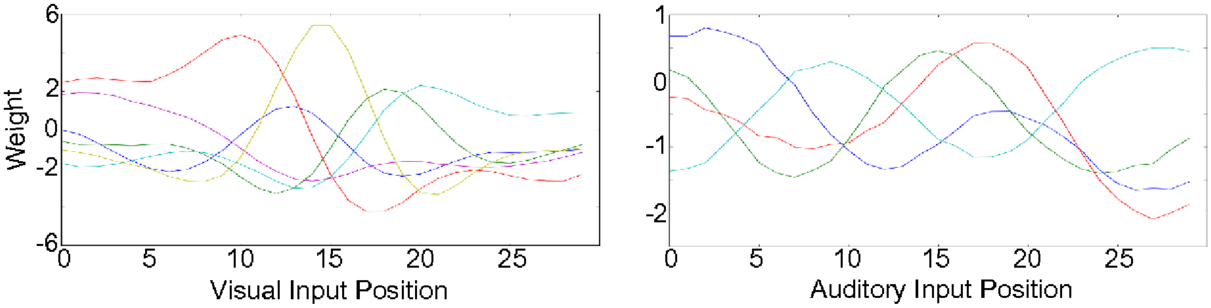

Figure 3.11. Input weights of the NN for the foveation setup. Input weights $v_{i,j}$ to representative hidden neurons. The left plot shows the weights only from visual input neurons ($i = [0 : x_{\max} - 1]$), the right only from the auditory input neurons ($i = [x_{\max} : 2x_{\max} - 1]$).

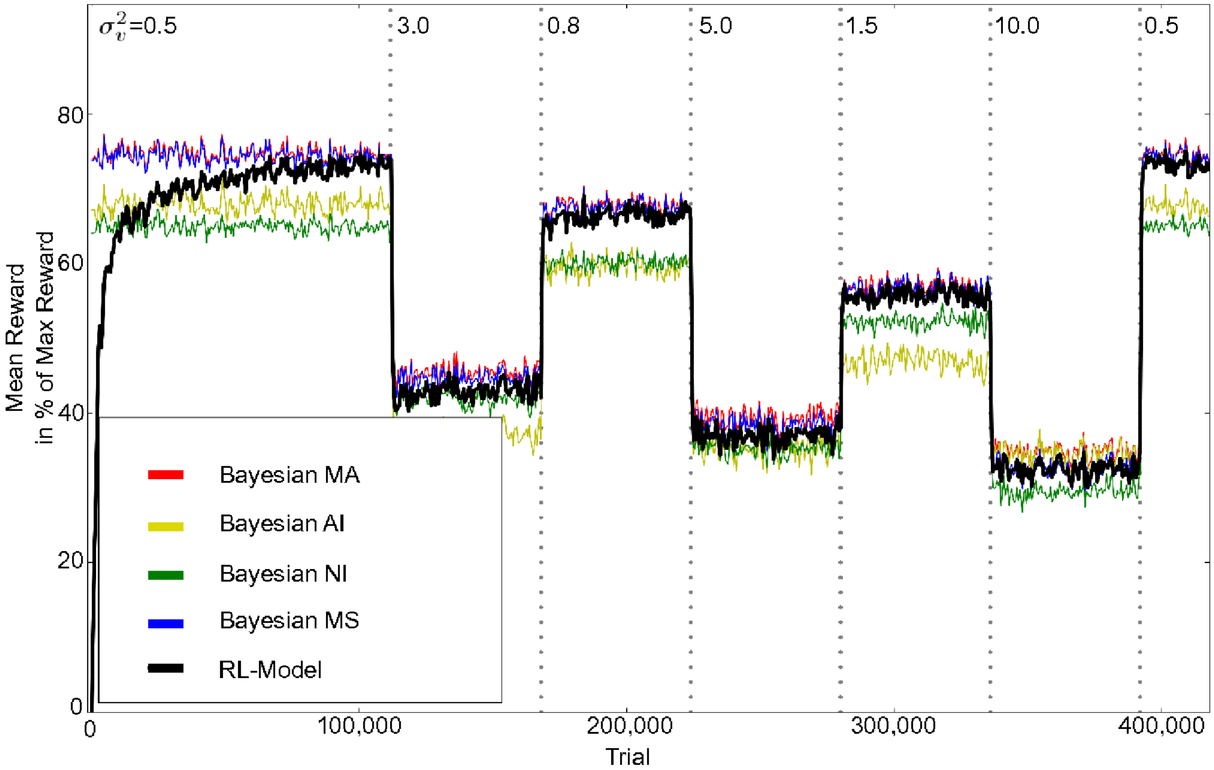

Figure 3.12. Performance of the RL Model for two outputs and temporally changing reliabilities. Reward obtained by the learner when choosing the action with highest predicted reward (black) compared to the different Bayesian observers. At each dotted vertical line visual reliability changes. Each data point is the sliding average of 1000 trials ($\sigma_a^2 = 3$).

distribution to a value different from zero.

Figure 3.13A shows contour plots of the Q-value function for one particular action after normal (filled) and biased training (empty). The area which favors integration (red) shifts by as many positions on the auditory axis as are introduced by the bias. The same is true for the unisensory tuning curves (Figure 3.13B), which were generated by plotting the response of the same output neuron to sequential single stimulation of each unisensory input neuron. These results are qualitatively similar to the ones reported by Wallace and Stein for the relationship between auditory and visual receptive fields of single neurons in cat superior colliculus.

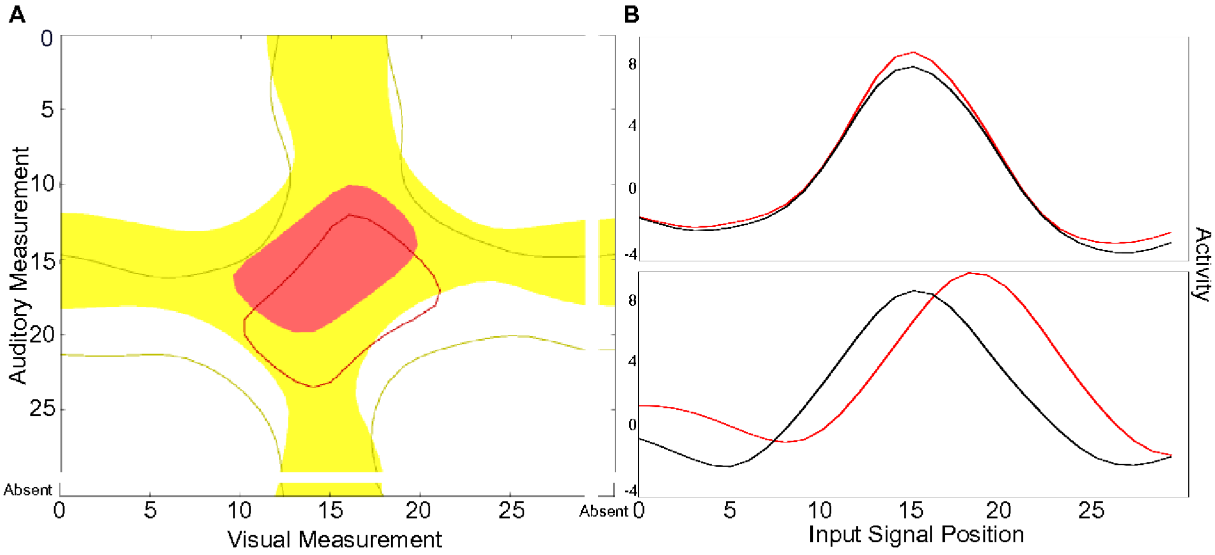

Figure 3.13. Responses of output neurons after training with auditory shift data. A: Overlay of contour plots of the Q-value functions for one action after unbiased training (filled areas) and after training with a 3-position shift in the mean of the auditory noise (empty areas). The contours include areas with predicted reward values higher than 10 (red) and 6 (yellow). B: Unisensory tuning curves of the same output neuron for biased (red) and unbiased conditions. The maximum visual response (top) does not change, whereas it is shifted by 3 positions in the auditory domain (bottom). ($\sigma_a^2 = \sigma_v^2 = 3$)

3.4 Discussion

The fact that cue integration in sensory inference can be well matched by Bayesian models has led to the suggestion that such computations are implemented in the brain by explicit computations with uncertainties. Accordingly, current research is looking for ways in which populations of neurons could implement Bayesian computations involving probability distributions [Fiser *et al.* 2010, Ma *et al.* 2006]. This view has led to the often implicit and sometimes explicit assumption, see e.g. [Körding *et al.* 2004, Knill & Pouget 2004, Whiteley & Sahani 2008], that reward dependent model-free learning can not mediate this behavior. Existing theoretical models look for structural features in neurons and networks that could produce the high integration performances. This leaves open, how cue integration and causal inference are learned over developmental timescales, as experiments with both children and animals suggest that cue integration abilities develop over time [Brandwein *et al.* 2011, Gori *et al.* 2008, Held *et al.* 2011, Nardini *et al.* 2008, Nardini *et al.* 2010, Neil *et al.* 2006, Putzar *et al.* 2007, Wallace & Stein 2007].

In this Chapter we investigated whether a reward-based model-free learner using function approximation is able to learn an orienting task requiring integration of cues. Furthermore, as the cues could either originate from different sources or a single one, it was necessary to learn when to combine the estimates, by taking into account that at larger separations of the two cues it is more likely that they originate from two different sources. The learner was given two audio-visual orienting tasks to solve. In the first task the learner was rewarded for orienting towards either one of the two stimuli, whereas in the second task the learner was rewarded separately for judging both the position of the visual and the auditory sources.

Under both task conditions the learner was able to carry out actions that combined cues according to their relative reliabilities. The reward obtained when following the reinforcement learner is higher than that obtained by the Bayesian learners that always or never integrate. It was also shown that the behavior of the RL model best matches that of a MA observer. This does not necessarily mean that humans always use MA, but shows the general ability of RL to approach optimal behavior. A recent paper by Wozny and colleagues [Wozny *et al.* 2010] found evidence for a majority of subjects acting most similar to PM (at the causal inference level), but also a significant number of people that were better fit by MA. A functional magnetic resonance imaging (fMRI) experiment on a slightly different task but also using two types of causal models found behaviour and BOLD activity favouring MA over MS [Wunderlich *et al.* 2011]. Further research is needed to clarify whether this is generally true or depends on additional parameters.

We could also show that the RL approach is able to deal with more complex uncertainty structures in the input. Here, the uncertainties are implicitly represented in the function approximation scheme of the value functions. Arguably, representing only uncertainties that are relevant for obtaining rewards is more economical than representing all potential distributions over all available scene variables. The distributions over sensory cues given relevant scene variables were not provided beforehand to the system, as is common in the Bayesian cue integration and causal inference setting. The proposed model was able to also perform similarly to the Bayesian predictions when the data likelihood was variable in time or space, when using non-uniform priors, and for changes in the causal structure. Humans were shown to be able to rapidly adapt to changes in cue reliability [Jacobs & Fine 1999, Triesch *et al.* 2002, Young *et al.* 1993] and causal layout [Wozny & Shams 2011]. Although we do not want to claim that this is necessarily mediated by reward for the very early adaptation, we show the potential of RL-mechanisms to react to those changes. It would also be interesting to test children for the developmental aspects of such rapid re-weighting [Bair *et al.* 2007], but more experiments will be needed to clarify those results.

One feature that is missing in our approach is temporal relations between signals, which in a natural environment provide an important cue for causal inference (e.g. [Lewald & Guski 2003]). It was shown that this influence is also plastic in children [Neil *et al.* 2006] and in adults [Fujisaki *et al.* 2004], so it would be interesting to see how reward mediated learning deals with the incorporation of temporal information. The TD-learning framework is in principle able to deal with delayed rewards. This question has to be addressed by future work.

All learning was done with immediate reward feedback to individual actions using learning rules that have been well established in conjunction with reward related learning and orienting movements [Schultz 2000,Schultz *et al.* 1997]. We are aware that using gradient descent learning to update the weights of the ANN could be considered problematic for a neural implementation [Crick 1989]. In the recent past, attempts were made to relate this kind of learning more closely to biology [D'Souza *et al.* 2010,Roelfsema & van Ooyen 2005, Tao *et al.* 2000]. In the Chapter 5 we will present an implementation that uses alternative solutions for learning of the synaptic weights.

Unfortunately we were not able to identify meaningful intermediate behavioral strategies while the model was still learning. It would be interesting to compare the behavior of the RL agent with recent empirical and theoretical work on the learning of cue integration, which suggest potentially different behavior such as calibration of a less reliable modality by a more reliable one [Gori *et al.* 2008, Gori *et al.* 2010,Knudsen & Brainard 1991,Strelnikov *et al.* 2011] or using the modalities alternatingly [Nardini *et al.* 2008] maybe according to the race model [Nardini *et al.* 2010, Neil *et al.* 2006]. The modality providing the basis for calibration could depend on the relative reliabilities, be innately determined or chosen at random. Consistent with the first option are results showing that even unisensory performance in certain non–visual tasks can be worse in early blind compared to sighted children [Gori *et al.* 2010].

To conclude, the RL algorithm with function approximation was capable of learning near optimal performance in the Bayesian sense for both cue integration and causal inference tasks (consistent with our results with tabular RL, see [Weisswange *et al.* 2009b]). Importantly, despite not performing explicit computations with uncertainties, the reinforcement learner successfully changed actions depending on the uncertainty in the stimulus. Considerable evidence about the neural basis of such algorithms makes this approach appealing. Furthermore, it gives a direct way of accommodating learning of cue integration and causal inference over developmental timescales. Thus, cue integration and causal inference can be done with performances close to optimal prediction using a model-free RL algorithm. In terms of the multiple controller hypothesis for behaviour (sec 2.3.3), we can say that, although there might be other parallel developments, the model-free learner could certainly play a role in the development of those inference abilities.

4

Application of reward-based learning of optimal cue integration to audio and visual depth estimation

4.1 Introduction

After showing that our reinforcement learning (RL) model is able to perform close to the Bayesian predictions in our well controlled simulation setting in Chapter 3, it is interesting to test its performance on real-world stimuli/tasks. From a different point of view, we also want to test its usefulness for applications in robotics and computer vision, where the integration of multiple sensory systems is often of interest [Hayman & Eklundh 2002, Khan & Shah 2001, Triesch & Eckes 1998, Triesch & von der Malsburg 2001]. Most real-world applications deal with large input spaces and can therefore not perform full Bayesian inference within reasonable time. Often an approximation is used, namely reliability-weighted averaging (see eq (2.6)). As was mentioned before, this approximation is fast to compute but will only be guaranteed to be correct in the case of independent (and Gaussian) noise on the cue estimates. In addition, its use requires the experimenter to know the likelihood variances for all cues. A fully autonomous robot though would be expected to learn these things and to be able to deal with many kinds of environmental dynamics. We therefore test if the reward-modulated learning system we proposed in the last Chapter could contribute to such autonomy, by learning to do cue integration without making assumptions about the environment.

We were able to run our model on two real datasets recorded with a robotic setup at Honda Research Institute Europe GmbH. These datasets consist of a number of cues from a depth estimation task and were in part published previously together with performance measures for each cue [Karaoguz *et al.* 2010, Rodemann 2010]. In these original publications the authors also tested cue combination with weighted averaging, but with mixed results. A few additional cues and measurements were added in the publication that is the basis for this Chapter [Karaoguz *et al.* 2011].

4.2 Methods & Original Datasets

4.2.1 Auditory Dataset

The fist dataset is of a binaural **auditory depth estimation** setup. Depth estimation for sound-sources is a common task in robotics, but usually done by triangulation either through a large array of microphones or using self-motion [Berglund & Sitte 2005, Nakadai *et al.* 2006, Sasaki *et al.* 2006]. To design more compact robots and to allow them to also estimate the depth of moving or short sounds (that would make motion triangulation much harder), it might be useful to find ways that are more similar to the biological implementations that usually use a number of cues from only two ears. We will introduce a number of those cues as well as the task setup in the following and then try to improve their performance by integrating them into a common estimate. For more details about the cues and the task setup also see [Karaoguz *et al.* 2011, Rodemann 2010].

Task Setup – Auditory Depth Estimation

The sounds to be localized were recorded by Tobias Rodemann in the robot lab of Honda Research Institute Europe (room dimensions: (12 x 11 x 2.8 m), echo: $T_{60} = 810$ms). 68 different sounds (speech, environmental sounds, music) were played from a loudspeaker set in front of the robot head. The distance between loudspeaker and head was varied in 9 steps using $[0.5m, 1m, 1.5m, 2m, 2.5m, 3m, 4m, 5m, 6m]$. Additionally, the head was rotated to 19 different pan positions for each depth. The database therefore consists of $19 * 68 = 1292$ samples at each distance. The recording set-up is sketched in Fig. 4.1 left.

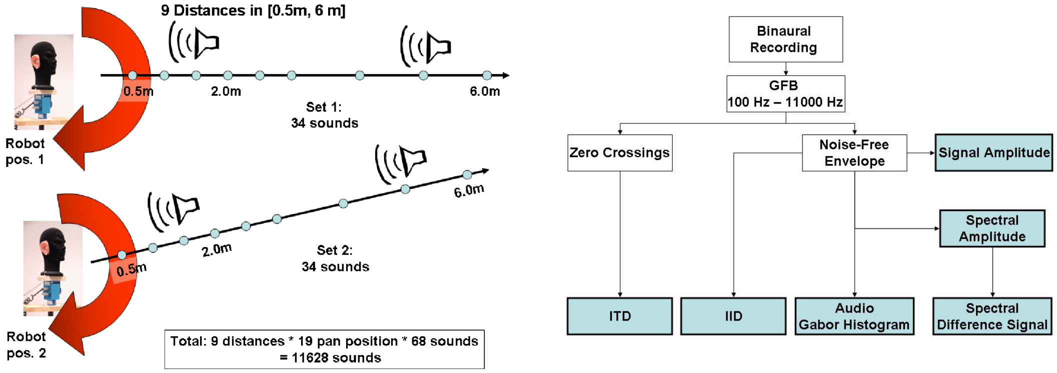

Figure 4.1. Experimental setup and processing diagram for the cues for the auditory depth estimation task Left: Experimental setup for recording the sound data. The loudspeaker was set at different distances from the microphones in the robot head, and the head itself was turned to generate different azimuth angles. 68 sounds were played in two sets recorded at slightly different positions of robot and speakers. Training and test data consists of half of the data from set 1 and set 2, each. **Right**: Basic flowchart of the processing system used to generate the responses of different cues. From these a depth estimate was computed by comparing each sample with an average response of the training stimuli at a certain depth. GFB stands for "Gamma Frequency Band". Images adapted from [Karaoguz *et al.* 2011] (©2011 IEEE).

Cue Descriptions – Auditory Depth Estimation

The following short explanations of each of the localization cues (see also Fig. 4.1 right) are taken from [Karaoguz *et al.* 2011], for more detail please see [Rodemann 2010]. A cue's response was taken to be the average over the full sound segment. Half of the sounds were used to generate a mean response profile of each cue for each depth (training). During testing each cue will produce a depth estimate by comparing the trained profiles at all depth with the current response and then picking the most similar one. The performance of each cue using these estimates is shown in Table 4.1.

Mean envelope amplitude
This cue represent the mean amplitude (related to mean energy of the sound) with roughly a $1/z$ relation to distance z. Since the measured signal amplitude also depends on the production amplitude (which is unknown), this cue depends on the distribution of production amplitude values in training and test datasets. The cue performed good for very close and far distances.

Spectral envelope
This cue measures the mean amplitude (energy) of the sound in different frequency bands. It is known that higher frequency bands are more strongly attenuated with distance than lower frequency bands. The performance was very weak over the depth range tested.

Binaural cues (IID and ITD)
Interaural Intensity Difference (IID) and Interaural Time Difference (ITD) are two standard cues for horizontal sound localization, showing a strong dependency on the azimuth angle of the sound relative to the robot's head. Both compute the difference of the incoming signal between the two ears, IID measures intensity, ITD the onset of the signal. They also exhibit a weak dependency on the sound's elevation (see [Rodemann *et al.* 2008]) and depth. These cues are quite useful especially at shorter depths, but decrease in performance when e.g. the robot is moving [Rodemann 2010].

Binaural spectral difference
Similar to IID and ITD this cue is based on differences in the signal recorded in two different microphones. However, while IID and ITD operate at single frequency channels, binaural spectral difference computes a histogram of binaural differences over a range of frequencies. This cue shows a similar performance as the above mentioned binaural cues.

Audio Gabors
For this cue histograms of filter responses are computed. These filters are 2D Gabor filters known from image processing, applied to the spectral envelope of the audio signal. This cue was not used in the literature before, but showed to have poor performance in this task.

Original Results – Auditory Depth Estimation

In the original publication Rodemann computed a number of quality measures for each of the cues, namely the mean localization error in meters, the relative error (mean error divided by distance), the probability of a mislocalization ("off-target"), and the probability of a severe mislocalization (near/far confusions), defined as instances where the estimated distance was more than 4 m away from the true depth. Table 4.1 summarizes the main results of the individual cues. Additionally he computed a multi-cue weighted average, with weights relative to the inverse variability of the cue estimates (see eq. (2.6)).

Figure 4.2 shows the estimation errors of the single cues for objects at different depths. The accuracy of most of the cues decreases with depth, only the spectral difference cue keeps its performance at all depths.

4.2.2 Visual Dataset

The second dataset from Karaoguz and colleagues consists of depth estimations from three visual cues [Karaoguz *et al.* 2010]. Visual depth estimation is probably the most common method to estimate distances in a 3D world, therefore much work in computer vision has been done on this task [Brown *et al.* 2003, Scharstein *et al.* 2002]. Most times two static cameras are used. In contrast, to record these data an active vision setup was used. The setup and the cues are presented in more detail in [Karaoguz *et al.* 2010].

4.2.3 Task Setup – Visual Depth Estimation

A robotic head with 2 degrees of freedom equipped with two cameras with an additional degree of freedom for each was used to record samples for the visual depth estimation task. The images had a resolution of

Table 4.1. Performance of auditory depth cues

Cue	mean error	rel. error	off-target	near/far conf.
Random	2.0	1.22	89%	9.5%
Amplitude	1.33	0.56	74%	2.8%
Spectral	1.71	0.98	78%	9.8%
IID	**0.46**	**0.15**	28%	1.5%
ITD	0.86	0.41	50%	2.3%
Gabor	1.82	0.78	85%	12.1%
Spectral Difference	0.52	0.32	**27%**	1.6%
Weighted Average	0.51	0.25	44%	**0.34%**

Combined results are computed as described in eq (2.6). For the combined cues the off-target value was computed by binning the estimated distance to the measurement distances. Table adapted from [Karaoguz *et al.* 2011] (©2011 IEEE).

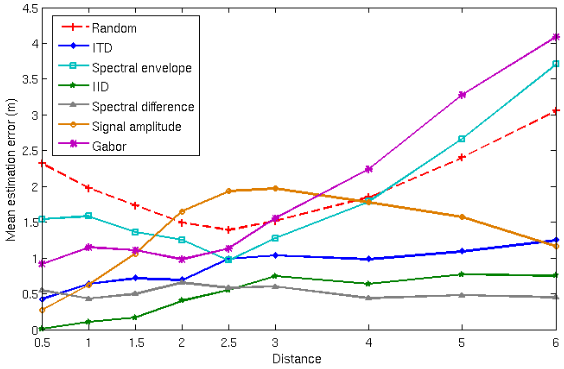

Figure 4.2. Spatial variability of error for auditory depth cues. Each curve shows the dependence of the estimation error of one of the auditory cues on the depth of the signal. Each datapoint represents the mean estimation error averaged over all sounds used at a certain depth. Figure first published in [Karaoguz *et al.* 2011] (©2011 IEEE).

400x300 pixels. A linear unit moves a small object platform along the depth dimension to autonomously generate data for a variety of depth values. In contrast to the auditory data, this set did not use fixed depth bins (using depths ranging from 300 to 1500mm). The procedure of alternatingly moving the platform and recording all cues was repeated for each of 11 objects from the HRI150 database [Kirstein *et al.* 2008]. One object was used for calibration purposes.

Cue Descriptions – Visual Depth Estimation

The description of the three visual cues for depth estimation are adapted from [Karaoguz *et al.* 2010, Karaoguz *et al.* 2011]:

Vergence
A visual system capable of changing its camera parameters can achieve stereo fixation on an object by positioning the intersection point of the line of sight of the two cameras on the surface of the object. The distance z to the fixation point can be derived from the triangulation using a pinhole camera model as:

$$z = \frac{b}{2\tan(\frac{\Theta_v}{2})}, \tag{4.1}$$

where Θ_v is the vergence angle and b is the baseline. The vergence angle was computed from left and right camera angles as $\Theta_v = \Theta_{left} + \Theta_{right}$. Vergence was assumed to be symmetric ($|\Theta_{\text{left}}| = |\Theta_{\text{right}}| = \Theta$).

Stereo Disparity
When two cameras are fixed to a point, visual stimuli in 3D space that are positioned in front or behind the camera's fixation point will be projected on to different locations in the images of the two cameras. This difference is referred to as stereo disparity. If the cameras can change their fixation angle, disparity values are relative to the current fixation point. Points belonging to an object under fixation have disparities of zero, those objects that are closer/further than the camera have negative/positive values. An active rectification method from [Dankers *et al.* 2004] was used to obtain absolute disparity values. After applying rectification, depth from disparity can be computed as:

$$z = \frac{bf}{d} + r + f, \tag{4.2}$$

where d is the distance between the left and right projection of the object, $f = 5.4$mm is the focal length and $r = 18.75$mm is the distance from the center of rotation of the cameras to the image planes. A block matching algorithm from OpenCV version 2.0 [Bradski & Kaehler 2008] was used to compute the disparity d from two images. To select the relevant disparity values that belong to the object a color based segmentation process was used.

Familiar Size
The perceived size of an object decreases with depth. If the true size of an object is known, this can be used to compute an estimate of the depth. Using a pinhole camera model that leads to:

$$z = \Big(\frac{fW}{w} + r + f\Big) cos(\Theta), \tag{4.3}$$

where $cos(\Theta)$ is close to 1 because of the small distance between the two cameras. The physical size W for all objects used in the experiments was measured, the retinal size w is computed using the same color based segmentation as for stereo disparity. The width of the objects was used for estimation because of better accuracy compared to the height.

Original Results – Visual Depth Estimation

Table 4.2 shows the mean estimation errors of individual cues and their combinations using weighted averaging in three ranges. The error is defined as the absolute value of the difference between the estimated depth and actual depth averaged over all objects. One can see that the combinations of cues did not produce the best results in all ranges. It seems that for these cues weighted averaging is not a very good approximation of Bayesian cue integration. One reason might be correlation in the noise between the cues due to similar preprocessing methods. Another one seems to be a dynamic bias for larger depth.

The change of accuracy over depth can be also seen in Fig. 4.3.

Table 4.2. Performance of visual depth cues

Cue	Near Dist.	Middle Dist.	Far Dist.
Vergence	**16.52** (6.65)	44.46 (13.18)	131.31 (46.99)
Familiar Size (FS)	44.64 (6.11)	86.09 (23.81)	175.38 (33.27)
Stereo Disparity (SD)	27.15 (13.91)	53.02 (21.13)	141.14 (32.47)
Combinations via weighted averaging			
Vergence+FS	17.63 (9.60)	**36.64** (14.55)	123.62 (109.52)
SD+FS	26.69 (19.60)	46.11 (29.03)	134.63 (82.24)
Vergence+SD	19.26 (12.85)	43.83 (16.60)	**120.73** (78.03)
Vergence+FS+SD	19.56 (13.68)	37.52 (19.47)	121.16 (77.92)

Mean (and standard deviation) of depth estimation errors (in mm) for single visual cues and combinations of them, averaged over all objects. The error values are shown for different depth ranges - Near: 300-700mm, Middle: 700-1100mm, Far: 1100-1500mm. Table adapted from [Karaoguz *et al.* 2011] (©2011 IEEE).

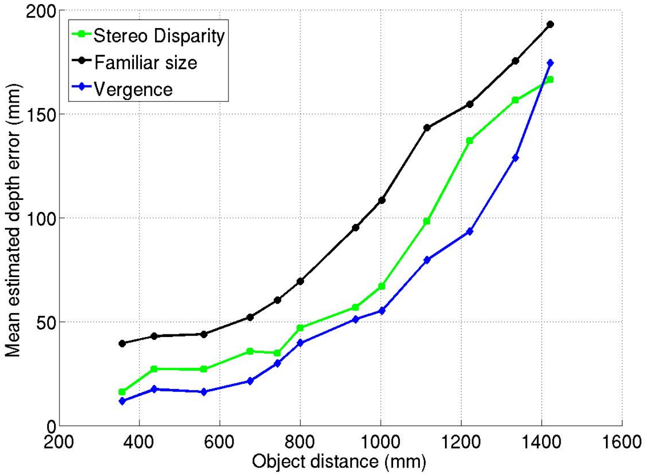

Figure 4.3. Spatial variability of error for visual depth cues. Each curve shows the dependence of the estimation error of one of the visual cues on the depth of the signal. Each datapoint represents the running average over 8 data points of the mean estimation error averaged over all objects at a certain depth. Figure first published in [Karaoguz *et al.* 2011] (©2011 IEEE).

4.2.4 Methods

The training of our RL-agent on the auditory task is done using 6 input populations (the number of cues) with 9 units each (the number of depths used) and $j = 30$ hidden units. One training sample consists of the depth estimate of each of the six cues for the given auditory signal encoded into a binary vector. The datasets provide the true depth for each sample, but we transform it into a reinforcement signal using as in the previous section

$$r(\hat{x}|X_{mathrmcue}) = \max(0, (\rho - \min(|X_{mathrmcue} - \hat{x}|)))^2 \tag{4.4}$$

with $\rho = 3$ and $X_{mathrmcue}$ being a vector with the estimates of all single cues. We use the stimuli from half of the objects for training, the other half for testing, as it was done to compute the weights and get the performance for the weighted averaging results. ν_ε and ν_τ were both set to $10,000$.

For the visual estimation task we setup our RL-agent using 3 input populations (the number of cues) with 245 units each (number of 10mm bins required to cover the full range of estimates of the cues) and $j = 100$ hidden units. Each training sample is encoded as one binary vector of length 245 for each cue with a 1 at the bin that contains the cues depth estimate for the current object. The true depth of each sample is also mapped to the bin-space, and then compared to the bin represented by the winning output unit to compute the reinforcement signal again using eq. (4.4) with $\rho = 15$. The split-up of the dataset for training and testing is done in a similar way as in the auditory task. We use $\nu_\varepsilon = \nu_\tau = 50,000$.

4.3 Results

4.3.1 Learning Auditory Depth Estimation

We train our RL-model using all the cues introduced above for the input. Figure 4.4 shows the results on the test sounds after 10,000 training steps compared to the performance of the best cue (IID) and to that of the weighted averaging approach.

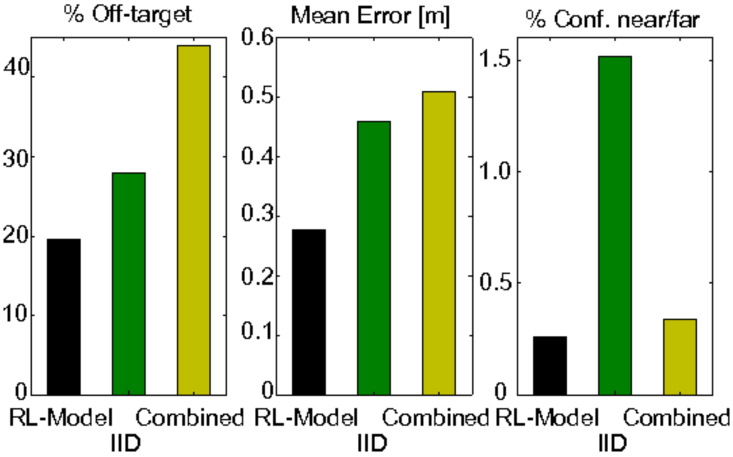

Figure 4.4. Performance of the RL-model for auditory distance estimation. The plots show from left to right the percentage of estimation errors, the mean estimation error in meters and the fraction of near/far confusions (errors of more than 4m). The RL-model's performance is compared to the best single cue (IID) and to a Bayesian weighted average of all 6 cues. Figure first published in [Karaoguz *et al.* 2011] (©2011 IEEE).

For all measures the model is better than the best single cue and better or equal to the weighted average of all cues. One reason for that can be seen when looking at the spatial change of the mean error for each single cue (see Fig 4.2). Some cues are for example very accurate at short distances but performance decreases with increasing depth. From that, one can predict that a single set of weights can not lead to an optimal integration at all depths. The neural network instead can learn to integrate the cues differently depending on the input pattern and thus performs almost equally well at all distances (Fig. 4.5).

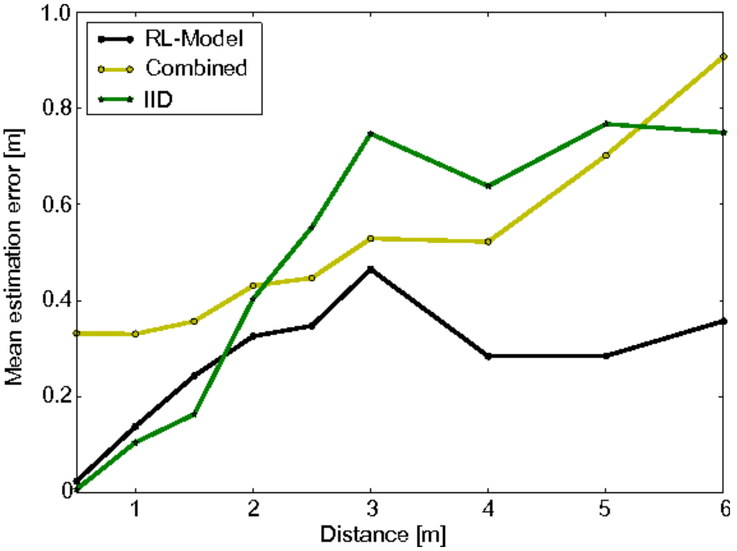

Figure 4.5. Spatial variation of auditory distance estimation error. Mean error for different depths of the weighted averaging (yellow) and the reward-based learning approach (black) for the auditory task in comparison with the best single cue (green). Figure first published in [Karaoguz *et al.* 2011] (©2011 IEEE).

4.3.2 Learning Visual Depth Estimation

For the visual depth estimation task we use a very similar setup, but the depth estimates of the single cues are continuous so that each neuron will now be active for inputs within a certain range. The output units have to represent discrete depth as well, therefore we use the same binning here. The RL-model's estimates can only be as accurate as allowed by the binning size. It is worth mentioning that the encoded bin size of the in- and output neurons does not necessarily have to be the same, but for simplicity we set all sizes to 10mm for the results shown here.

As can be seen from Fig. 4.3 and Table 4.2 the three cues change in quality relative to each other, similar to the cues in the auditory task. It has been explained that estimation errors from individual methods can be reduced by improving the accuracy of the visual system however, trends will stay the same [Karaoguz *et al.* 2010]. Therefore, it will not affect the cue integration process. The error after weighted averaging of multiple cues still has a strong tendency to increase with distance (Fig. 4.6 yellow curve). This supports the notion of the cues showing a depth dependent bias. Finally we can also expect correlations in the noise distribution of different cues, since for example both stereo disparity and familiar size use the same segmentation method. Figure 4.6 plots the performance of the neural network after 100, 000 training steps as a function of depth. We transformed the error to metric distances (from an error as a number of bins) by computing the distance of the true value to the center of the depths range represented by each bin. As in the auditory task, we get an error smaller or equal to both best cue estimate and weighted averaging, with a much lower increase with depth.

This can also be seen if we separately compute the errors for groups of near, middle and far distance to compare it with the results shown in Table 4.2. The RL-model produces mean errors of 16.6, 34.6, and 60.5mm respectively. For comparison, the best values for each distance among single cues or weighted averaging were 16.52, 36.64, and 120.73mm (see table 4.2).

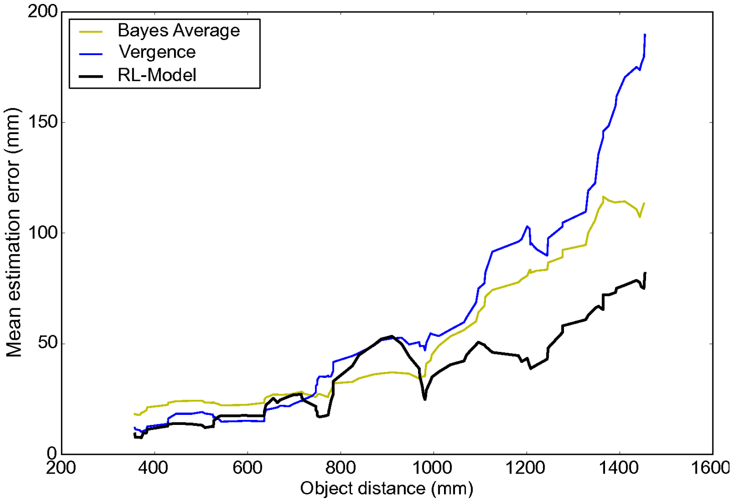

Figure 4.6. Spatial variation of visual distance estimation error. Mean error for different depths of the reward-based learning approach, the Bayesian average and the best single cue (vergence) for the visual task. Mean values of 10 repetitions of each leave-one-out training trial. Plot shows the running average over 8 data points. Figure first published in [Karaoguz *et al.* 2011] (©2011 IEEE).

4.4 Discussion

We show that the reward-modulated cue integration approach can work with more natural data in a real world experimental setting. In particular, it can handle not only artificially imposed noise but also signal distortions that originate in the environment, like different sounds or azimuth changes. Similar to the simulation results with spatial variation in cue uncertainties, the learner adapted to variations in the relative reliabilities of the depth cues over the input space. These results support our proposal that reward-mediated learning could contribute to the development of cue integration abilities in humans.

The performance of our network does not only match but often even exceeds that of a weighted averaging approach. Importantly, we used the same number of training stimuli as were used to determine the Bayesian weights of all the cues. The main computational load of the system is needed during training, while generating a single depth estimate is only slightly slower than using the weighted averaging (it essentially requires two matrix multiplications). The fact that the model performed better than the common approximation of Bayesian integration shows that the necessary assumptions for the latter might not always be met in natural tasks. Learning the integration strategy can for example better balance sensory correlations, like those potentially introduced by using the same pre-processing for multiple cues in the visual task.

Using RL based mechanisms can make training more flexible and independent of the availability of labeled data. The reward signal we used for the depth datasets was computed based on the true depth, which is known from the way the data was generated. The quality of an action though could also be measured in many different ways. One example for the depth estimation tasks would be the success of a grasping movement based on these estimations. If such an online reward signal is available, the model could adapt the integration even during the operation, assuming that tasks are executed frequently. Alternatively the true value could be provided by a precise but computationally costly cue, which after some training can be replaced by the integration of a group of cheaper cues.

In general, integrating information from multiple cues has great potential to improve the performance of an agent in many different tasks. Having a general and fast mechanism for the combination of those cues is therefore beneficial for many applications in robotics and computer vision. The reward-mediated learning algorithm presented in this thesis is shown to be a promising alternative over common approximation methods for Bayesian inference, especially in terms of autonomously learning robots.

5 Learning Cue Integration in a Reservoir Network

5.1 Introduction

In the previous Chapter we showed that the concept of reward-mediated learning can lead to the development of behaviour with performance close to the Bayesian predictions. Here we want to demonstrate that this principle can also be transported on to a system that can in more detail address the *implementational* level of the brain. Particularly, we show that the combination of a number of biologically plausible learning mechanisms can drive the network towards this kind of solution, which can not be obtained by either of these mechanisms alone. I will first provide some more background about plasticity mechanisms working at the single cell level. There is a great number of mechanisms both found experimentally and proposed theoretically, where the latter often tries to explain biological phenomena. I will focus only on those theoretical formulations that are both experimentally well supported and of relevance for the use in our model. Our model is a recurrent neural network inspired by liquid state machines and echo state networks (or "reservoir computing"), which will also be introduced in the following. The difference to those original proposals is (I) the use of very simple binary threshold neurons and (II) the use of a variety of plasticity mechanisms within the reservoir. We decided to use this method because recent work from our lab could show that despite its simplicity it can develop many features that also seem useful for our task [Lazar *et al.* 2009,Savin & Triesch 2010]. Specifically, the basic implementation is adapted from that used in [Savin 2010].

5.1.1 Neural Plasticity

Synaptic Plasticity

The most influential theory about plasticity at neuronal synapses goes back to Donald Hebb [Hebb 1949]. In his book he introduced the theory that neurons that spike at a similar point in time will increase their connection strength with each other, "neurons that fire together, wire together" (this is now referred to as Hebbian learning). Later the group of Kandel and others could show that those principles are indeed present at neuronal connections (e.g. [Hawkins *et al.* 1983, Wigström & Gustafsson 1986]). Generally, there seem to be two main mechanisms, one that is increasing synaptic strength and is termed long term potentiation (LTP) [Bliss & Lø mo 1973], the other having the opposite effect, called long term depression (LTD) [Lynch *et al.* 1977]. More recently theorists proposed a refinement to Hebb's law, which emphasizes the causality between the two firing events in different neurons by including temporal order. In so-called spike-time-dependent plasticity (STDP) [Gerstner *et al.* 1996, Song *et al.* 2000] a synapse is potentiated if the pre-synaptic neuron spikes shortly before the post-synaptic neuron, but will be depressed for the reverse order (Fig. 5.1). Around the same time experimental results on the influence of timing in plasticity showed evidence that this STDP is indeed present in the brain [Bi & Poo 1998, Bi & Poo 2001, Dan & Poo 2004, Froemke & Dan 2002, Markram *et al.* 1995, Markram *et al.* 1997, Zhang *et al.* 1998]. The

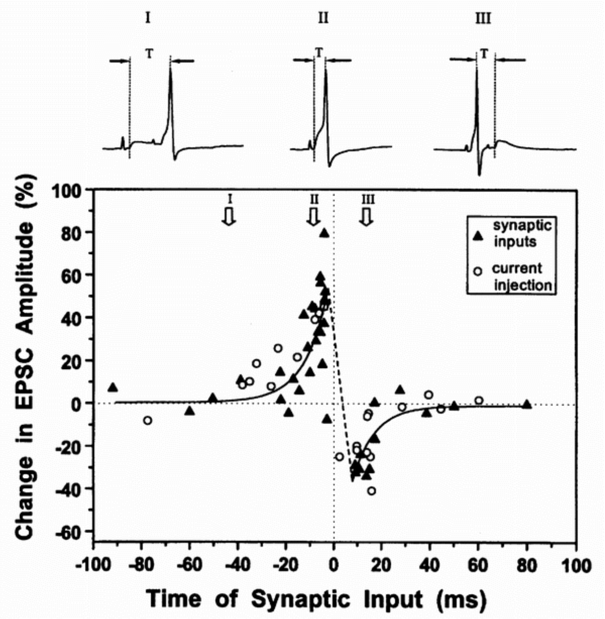

Figure 5.1. Spike-time-dependent plasticity The plot shows the influence of the time difference between a spike in the pre-synaptic neuron and a spike in the post-synaptic neuron on synaptic plasticity. If the pre-synaptic neuron fires shortly before the post-synaptic neuron (II) LTP is induced with its amplitude decreasing with longer delays between the spikes (I). For the inverse order (III) one can find LTD, again stronger for spikes closer in time. The synaptic weight change is measured as the change in amplitude of excitatory post-synaptic potentials. Data points are recorded in tectal neurons of frog *in vivo*. Figure reprinted by permission from Macmillan Publishers Ltd: Nature ([Zhang *et al.* 1998], ©1998).

underlying mechanism is based on the backpropagation of action potentials from the post-synaptic neuron's soma into the synapse using Calcium (Ca^{2+}) signals. NMDA receptors (NMDAR) are thought to detect coincidences between those Ca^{2+}-signals and the binding of neuro-transmitters released from the pre-synaptic site. In many studies since it was shown that STDP, despite varying in parameters like the size of the LTP and/or LTD window, can be considered one general principle of neural plasticity [Caporale & Dan 2008].

Most of the work on synaptic plasticity, both experimental and theoretical, is focusing on excitatory synapses. Nevertheless there is also evidence for adaption at inhibitory synapses (see reviews in [Castillo *et al.* 2011, Feldman 2009, Kullmann & Lamsa 2007, Maffei 2011]. In many cases similar mechanisms to those found at excitatory connections can be shown to also exist at inhibitory synapses (e.g. LTP [Grunze *et al.* 1996]). Only recently theoreticians also started to research that matter.

One proposal by Bourjaily and Miller for example [Bourjaily & Miller 2011b, Bourjaily & Miller 2011a] which is based on experimental findings in [Maffei *et al.* 2006] proposes a mechanism that increases the synaptic weight from an inhibitory to an excitatory neuron when the former fires and the latter is depolarized but silent. This rule was termed long term potentiation of inhibition (LTPi) and shown in simulation to increase the selectivity to stimulus combinations in excitatory neurons leading to an improved performance of a network in an XOR logic task.

A second mechanism proposed by the group of Gerstner [Sprekeler *et al.* 2011, Vogels *et al.* 2011] is based on data from [Dorrn *et al.* 2010], where it was shown that the correlation between inhibitory and excitatory activity increases with experience. From that a plasticity rule was constructed that potentiates inhibitory synapses if it spikes close in time (ignoring the spike order → Hebbian type) with the excitatory neuron, and slightly depress it for every other pre-synaptic spike. An analysis showed that this plasticity can shape the balance between excitation and inhibition in a network of spiking neurons.

Despite these recent studies, there is still relatively little knowledge about the interactions of inhibitory plasticity with other forms of neural learning.

Reward-modulated Plasticity

As was discussed in Section 2.3, the neuromodulator dopamine (DA) is thought to be the main factor in the biological implementation of reinforcement learning (RL). Some DA-releasing neurons seem to encode a reward prediction error [Schultz *et al.* 1997] and project to large parts of the brain [Jay 2003, Kandel *et al.* 2000]. In many synapses dopamine receptors sit close to NMDA and AMPA receptors (AMPAR) but do not directly influence the post-synaptic potential. Two main types of receptors can be found (D1 and D2) which are sometimes expressed together in the same neuron, sometimes selectively expressed in certain populations [Creese *et al.* 1983]. These receptor types differ in their sensitivity to spatial and temporal factors of dopamine in the medium [Sealfon & Olanow 2000] and can elicit different effects when activated [Surmeier *et al.* 2007]. This has lead to a proposal that they might represent two distinct mechanisms of synaptic modulation [Bromberg-Martin *et al.* 2010, Shen *et al.* 2008].

These results prompted for a more detailed investigation of the influence of DA on synaptic plasticity which would allow rewards to impact behaviour. In an early study on LTD in striatal slices from mice Calabresi and colleagues measured the impact of a variety of substances on plasticity [Calabresi *et al.* 1992]. One of their findings was that blocking DA receptors or removing the DA from the bath impaired LTD. Interestingly it seemed that the activation of both DA receptor types was necessary to allow tetanic stimulation to depress a synapse.

In a different study in vivo, the pairing of an auditory signal with the activation of dopamine neurons increased the area in auditory cortex that was tuned to that signal [Bao *et al.* 2001]. Similarly the pressing of a lever that directly stimulated DA neurons in rats lead to potentiation of the motor connections active at lever press [Reynolds *et al.* 2001]. The effects on synaptic plasticity can vary from lowering the threshold for induction [Lemon & Manahan-Vaughan 2006] to be in a necessary condition to enable STDP [Pawlak *et al.* 2010]. In one study it was shown that D1 receptor activation is an absolutely necessary requirement, but D2 activation is only modulating the shape of the plasticity window [Pawlak & Kerr 2008]. In general there is more and more evidence that in many parts of the brain DA is directly controlling synaptic plasticity [Calabresi *et al.* 2007, Jay 2003].

Recently theorists have also started to incorporate dopamine signals in synaptic learning rules [Farries & Fairhall 2007, Reynolds & Wickens 2002, Wickens *et al.* 2003]. The biggest progress was made when those rules were derived analytically and were able to solve the problem of the sometimes large temporal delay between neuronal co-activation and the reward signal [Florian 2007, Izhikevich 2007, Xie & Seung 2004]. In [Izhikevich 2007] a so called eligibility trace was introduced which stores a potential for a weight change based on classic STDP. This trace is decaying exponentially with time and is used whenever a reward signal is delivered to change the synaptic weight. This idea is called reward-modulated spike-time-dependent plasticity (R-STDP). A similar proposal in [Florian 2007] computed the potential weight change using a rule analytically derived from partially observable Markov decision processes [Bartlett & Baxter 2000] instead of STDP. It was shown that the latter rule is more stable when using the prediction error instead of the absolute reward, as proposed by the former, to modulate the plasticity [Frémaux *et al.* 2010]. Simulation and analysis results from these and other papers [Legenstein *et al.* 2008, Vasilaki *et al.* 2009] demonstrated that networks using R-STDP are able to learn complex tasks.

Homeostatic Plasticity

Besides synaptic learning in reaction to concrete input instances, neurons were also shown to be able to adapt to long lasting changes in the statistics of these inputs. If the activity of neurons in a slice was blocked for an extended period and afterwards tested and compared with that before the alteration, experimenters did find an overall increase in firing rates [Desai *et al.* 1999b, Turrigiano *et al.* 1998, Turrigiano & Nelson 2004, Turrigiano 2011]. When chemically increasing the firing rates the reverse was true. Those results could later be replicated in vivo [Pratt & Aizenman 2007].

It could be shown that there are actually two different mechanisms involved in this effect. Those mechanisms are called homeostatic plasticities because they aim at keeping the activity of a neuron within a stable range. The first mechanism, often referred to as *intrinsic plasticity (IP)*, changes the intrinsic excitability of the neuron by adapting the firing threshold, or more general the transfer function between membrane potential and spiking [Desai *et al.* 1999b, Desai *et al.* 1999a, Paz *et al.* 2009].

A theoretical interpretation for the function of these changes is the optimization of information transfer given some limitations [Stemmler & Koch 1999]. The firing rate, which is one major way of information transmission for a neuron is limited to values between zero and a maximum determined by the temporal extent of an action potential. For an optimal use of this limited range every firing rate should be used with the same frequency. But if one also takes into account the energy cost for spiking, it should be tried to keep the mean firing rate at a small value [Laughlin *et al.* 1998]. This would lead to an exponential distribution of firing frequency, which is indeed what was found in responses of neurons in monkeys watching natural videos [Baddeley *et al.* 1997]. Triesch derived learning rules along the same lines for more abstract neuron models [Triesch 2005b], including one for online adaptation of the neuron's transfer function [Triesch 2005a].

The second mechanism is thought to balance the synaptic changes imposed through Hebbian learning, which is known to lead to run-away potentiation of some of the weights. This is intuitive, since Hebbian learning strengthens synapses between correlated neurons, this strengthening in return will enhance the correlations and so on. One possibility is to adjust the synaptic learning rule by adding a term that takes care of this problem. In [van Rossum *et al.* 2000] for example it is proposed to scale the potentiation amplitude of an STDP rule with a value inversely related to the current weight of a synapse. The rule for depression is the same for all synapses. This is can be matched to the biological finding that strong synapses change much less compared with weak synapses [Bi & Poo 1998].

Alternatively all synapses can be scaled by a single value. Experiments suggest that this value is multiplicative and depends on the overall activity of the neuron [Shepherd *et al.* 2006, Turrigiano *et al.* 1998, Turrigiano 2008]. Similar to the experiments for IP, suppressing all activity in a brain slice leads to a later increase in firing activity in all neurons, which here could be attributed to an upscaling of synaptic weights. Interestingly such scaling is not confined to excitatory synapses, but seems to be also present at inhibitory synapses using a different scaling factor [Hartman *et al.* 2006, Kilman *et al.* 2002].

Computationally such *synaptic scaling* is appealing because it keeps the relative weight difference constant, and at the same time imposes a form of competition between pre-synaptic neurons. The idea of multiplicative scaling is therefore present in the field since the 1970s time [von der Malsburg 1973].

The interaction between these two (or more) homeostatic mechanisms is in most part still to be worked out [Nelson & Turrigiano 2008, Turrigiano 2011]. Interestingly, the BCM rule [Bienenstock *et al.* 1982], one famous theoretical mechanism developed to prevent run-away excitation in Hebbian learning, could be shown to be matched in some cases by a combination of STDP and IP [Savin *et al.* 2010].

5.1.2 Reservoir Computing

The classical artificial neural network (ANN) has only feed-forward connectivity and is not able to approximate functions that have a temporal component. To incorporate a form of memory one can add recurrent connections, that is connections within the same layer. Such recurrent neural networks (RNNs) were proven to have the potential of being universal approximators of dynamical systems and non-linear functions [Funahashi & Nakamura 1993]. Training those networks becomes much more complicated though. Classical gradient based methods can not be guaranteed to converge anymore [Werbos 1990], and the temporal unfolding of the network to update the recurrent weights is computationally expensive [Williams & Zipser 1989]. Some more recent learning methods started to separate the training of the output from that of the hidden layers to improve speed and performance [Atiya & Parlos 2000, Schiller & Steil 2005, Steil 2004].

This line of thinking finally resulted in the development of reservoir networks that use randomly connected recurrent hidden layers (the "reservoir") and train only the output weights (for overviews see [Lukoševičius & Jaeger 2009, Verstraeten *et al.* 2007]). Two versions of these networks were originally proposed independently of each other: Echo state networks (ESNs) [Jaeger 2001] came from a machine learning background, whereas liquid-state machines (LSMs) [Maass *et al.* 2002] were developed as a computational neuroscience method. The reservoir neurons of an ESNs are often sigmoidal or leaky integrator neurons. LSMs use more biological models like leaky integrate-and-fire (LIF) neurons [Gerstner & Kistler 2002], which makes it usually more difficult to implement and fine tune the parameters. It was shown that ESNs can vastly outperform previous systems for tasks like chaotic systems prediction [Jaeger & Haas 2004] and that in general reservoir computing is computationally universal for continuous-time, continuous-value real-time systems modeled with bounded resources [Maass *et al.* 2003, Maass *et al.* 2006]. The output weights are usually learned offline in a supervised fashion using standard linear regression methods[1].

Both systems are very sensitive to a variety of reservoir parameters, e.g. connectivity strength or network size [Lukoševičius & Jaeger 2009]. Only few rules of thumb for choosing those parameter values are known by now, e.g. that the largest eigenvalue of the reservoir weight matrix should be smaller than one [Jaeger 2001] (but see e.g. [Steil 2007]). Mostly large reservoirs with sparse and random connectivity are used together with a dense input and output connectivity [Jaeger 2001]. Interestingly using LSMs with parameters based on statistical properties of the brain were shown to perform very well [Haeusler & Maass 2007]. Another possibility is to use evolutionary methods to optimize reservoirs based on the performance of the linear read-out units (e.g. [Schmidhuber *et al.* 2007]). In an attempt to model Bayesian computations with recurrent networks, Rao [Rao 2004] came up with a formal description of the features that need to be encoded in the recurrent weights to do probabilistic inference. Based on this description he then used an approximation and solved the resulting equation using pseudo-inverse matrices. This method is completely supervised, meaning it uses the full knowledge over the true structure, likelihoods and so on for training. He later proposed a similar model that in addition uses RL, again encoding Bayesian principles explicitly in the parameters [Rao 2010].

Originally reservoir computing was defined not using any (online-)learning in the reservoir[2], but recent research does more and more also look at (unsupervised) plasticity mechanisms that could help finding "good" recurrent weights. Pure Hebbian or Anti-Hebbian rules did not improve performance [Jaeger 2005]. In contrast STDP could improve input separation for natural speech stimuli [Norton & Ventura 2006], and a combination of STDP in the reservoir and a supervised mechanism that adapts spike transmission delay timing in the output connections improved classification of temporally encoded inputs [Paugam-Moisy *et al.* 2008]. Similarly using IP in the reservoir neurons can also improve the

[1] But there is also work on alternative schedules like e.g. evolutionary algorithms [Xu *et al.* 2005].

[2] It should be mentioned that most LSMs were using short-time plasticity, which can depress or facilitate the transmission at a single synapse during the arrival of high frequency stimulation [Maass *et al.* 2002], a mechanism known to occur at real synapses [Markram *et al.* 1998]

performance of a network [Schrauwen *et al.* 2008, Steil 2007].

After finding evidence for the beneficial influence of plasticity mechanisms in the reservoir the logical next step was to test if the improvements using different methods are redundant or could actually add up when used together. This is particularly interesting since it was shown before that in single neurons the combination of IP with Hebbian learning [Triesch 2005b, Triesch 2007], or STDP [Savin *et al.* 2010], enables the neuron to discover heavy-tailed distributions in the input and is able to perform computations similar to independent component analysis (ICA), that is doing blind source separation [Hyvärinen *et al.* 2001]. Lazar and colleagues introduced a reservoir network with binary threshold units, that uses IP and STDP to adapt the reservoir weights [Lazar *et al.* 2007]. Additionally they use a k-winner-take-all (WTA) mechanism to induce competition amongst the reservoir neurons. They did find an increase in stable limit cycles in the network state dynamics when comparing the dual plastic network with those only using one or the other or no plasticity mechanism. The performance on predictions of Markov processes did also increase, mediated by the development of a fading memory within the reservoir (note that the neurons themselves are memoryless). The same group later improved their model by replacing the k-WTA mechanism by a synaptic scaling rule [Lazar *et al.* 2009]. This new network again outperformed static reservoirs on a "counting" task on time-series with repetitive symbols and it was shown that removing any of the three plasticity mechanisms results in a decay in performance. In a later publication it was also demonstrated that the weights of the network can be interpreted as encoding a form of input prior, which leads to a spontaneous activity (no input to the reservoir) showing similar statistics to those elicited by the training inputs [Lazar *et al.* 2011].

Another network similar to this latter one introduced the use of reward-modulated plasticity within the network with the goal to not only optimize it for the input statistics, but also to e.g. better separate inputs for which classification by the output layer is more difficult [Savin & Triesch 2010, Savin 2010]. Specifically, the network uses R-STDP with eligibility traces from [Izhikevich 2007] in combination with IP and synaptic scaling. In contrast to almost all work on reservoir computing, including those mentioned above, in which the output is trained in a supervised and offline procedure, here it was necessary to get a reward signal based on the output signal in every trial. In an attempt to also have a biologically plausible training of the output weights, the same R-STDP rules as in the reservoir were used again combined with IP. Importantly the learning rates of those output adaptations were set to be much higher than those of the reservoir, to allow the output units to quickly adjust to plasticity induced changes in its inputs. The authors showed that such a setup can solve a delayed classification task, where an input stimulus is given at the beginning of a trial and after running the reservoir for a number of timesteps the output units had to estimate the class of this stimulus. Even more than in the previous work with binary threshold neurons a stable memory arose, which was much weaker or absent when leaving out any of the plasticity mechanisms or when replacing R-STDP with STDP. This work also showed that online output learning with similar plasticity rules is able to perform well, something that was already suggested in earlier analytical work [Legenstein *et al.* 2008].

Our work will try to extend the work on the interplay of plasticity mechanisms in the same type of simple networks. We will demonstrate that such a plastic reservoir in combination with an online spike-based output training is able to integrate noisy cues in a sensible way, in contrast to purely random networks.

5.2 Methods

We use the same bimodal orienting task as in Chapter 3 with two temporal variations introduced below. The reservoir of the recurrent network we use consists of excitatory and inhibitory binary threshold neurons. The connectivity within the excitatory pool is sparse with 10% of the possible connections present, between excitatory and inhibitory neurons exists a connectivity of 25% in both directions. The former synapses are adapted during training using binary STDP in combination with multiplicative synaptic scaling. Projections from inhibitory to excitatory neurons exhibit inhibitory STDP similar to [Sprekeler *et al.* 2011, Vogels *et al.* 2011]. Finally a binary threshold IP rule [Lazar *et al.* 2007] enforces a low average firing rate of the excitatory neurons.

Distinct sub-populations within the reservoir receive external input, where each one of those sub-populations gets activated by exactly one stimulus position in one modality (see Fig. 5.3 for details). The two different task schedules we use differ in the number of timesteps at which such an input stimulus is provided. In the **delayed response task**, the network only receives external activation in the beginning of a trial and can then run on its own for a number of timesteps. In the **accumulation task** in contrast, at every timepoint a stimulus is provided and therefore the network permanently gets external input. In both cases the network's output layer activity is evaluated after a certain delay dT. This output layer has full connectivity with the excitatory neurons of the reservoir and also consists of binary threshold neurons. Additionally output neurons compete via a WTA mechanism that only allows the one unit to fire that has highest activity above threshold. The output weights are trained in parallel with the reservoir (that is "online") using R-STDP [Izhikevich 2007]. The modulating reward signal is the difference between the predictions of a critic (external to the recurrent network) and the true reward received based on the current output activity.

5.2.1 Task

The general task setup is the same as in Chapter 3, Fig. 3.1. An agent receives two stimuli z^a and z^v from different modalities (for convenience again called the auditory and the visual modality), which provide noisy information about the position z^* of an object, randomly drawn from a uniform distribution over the space $|Z| = 20$ in each trial. The noise is additive, discretized and Gaussian distributed with variances σ_a and σ_v. In the causal inference setup we randomly choose if there is one multisensory ($C = 1$) or two unisensory ($C = 2$) objects. In the latter case the true positions z_a^* and z_v^* are drawn independently, in the former case we write $z_a^* = z_v^*$. The task for the agent is to correctly estimate and orient towards either of the true object positions. Depending on the distance $\Delta_{error} = \min(|z_a^* - z'|, |z_v^* - z'|)$ between the estimated position z' and the closest true position, a reward is given based on:

$$r(t+1) = \frac{((\rho - \Delta_{error})^2)}{r_{\max}}(\Delta_{error} \leq \rho). \tag{5.1}$$

ρ is a parameter that determines the error size that is still positively rewarded.

The main difference to the task in Chapter 3 is the simulation of a temporal task profile. The two variants we use are shown in Fig. 5.2. In both versions, the agent has to respond with a delay dT after the first stimulus is presented. The network is always simulated for $d_{Trial} = 5$ timesteps for each trial, independent of dT (but therefore limiting $dT \leq 3$)[3].

In the version shown on the left of Fig 5.2 only in the first step a stimulus is provided, and after a waiting time a response is required. We call this the **delayed response task**, which can also be seen as a multisensory version of the task used in [Savin 2010]. It can also include an additional delay dt_{av} between the stimuli in both modalities. $dt_{av} = 1$ for example would mean that the auditory stimulus is given one timestep after the visual one, $dt_{av} = -1$ would be the same but with the reverse order of the modalities. The idea behind this is to use the timing of the stimuli as potential additional cue for causal inference, by e.g. setting $dt_{av}^{C=1} = 1$ for a single cause and draw $dt_{av}^{C=2}$ uniformly from $\{-1, 0, 1\}$.

The second variation is inspired by models of decision making, that assume that evidence for each choice is accumulated over time [Usher & McClelland 2001] (like e.g the drift diffusion model (DDM)).

[3] Since the performance was usually decreasing with increasing dT, we did not intensively test the network with bigger values. Simulations with $d_{Trial} = 10$ for various small values of dT did not influence the results. A test for $dT = 5, d_{Trial} = 10$ did confirm the tendency for a decrease in performance (data not shown)

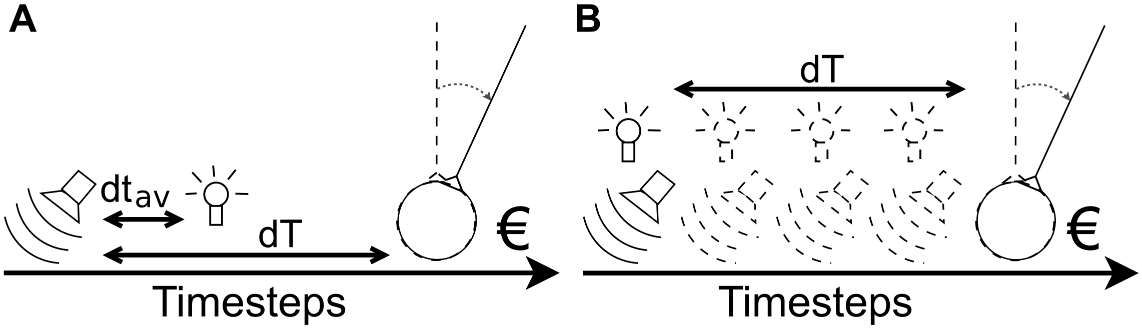

Figure 5.2. Temporal profile of the orienting tasks. **A**: The delayed response task profile. One of the stimuli is shown at the start of the trial and the second one follows with some delay dt_{av}. Which modality is stimulated first depends on the sign of dt_{av}. The agent's response is read out with a time delay dT after the first stimulus, and a reward is delivered. **B**: In the accumulation task profile, dt_{av} is kept at 0 and the agent receives both inputs at each timestep during dT. Information from each timestep can be accumulated to improve the final orienting movement after dT.

We will therefore call this variant the **accumulation task**. If an object is presented for a longer time, a sensory receptor will be able to acquire the corresponding stimulus multiple times. For internal noise it is proposed, that at each timestep the system receives an independent sample from the probability distribution over the relevant variable. Therefore, integrating all these samples into a posterior $p(z_a^*|z_{t_1}^a, ..., z_{t_n}^a)$ will improve the final result over that obtained when considering only a single stimulus, i.e. using $p(z_a^*|z_{t_1}^a)$. The temporal setup for this task is shown in Fig. 5.2 right, in every one of dT timesteps both a auditory and a visual stimulus is shown. The true position and number of objects stays constant throughout the trial, but every stimulus uses an independent sample from its modality's noise distribution. We will compare the performance of the agent to models only using the last stimulus for their estimates and to observers that accumulate information over dT.

5.2.2 Network Setup

A sketch of the reservoir network that we use is shown in Fig. 5.3. It consists of two populations of neurons, one excitatory and one inhibitory (in a ratio of 4:1, as was found in neuroanatomical studies [Gabbott & Somogyi 1986]). We use sizes of $N_e = 600$ excitatory and $N_i = 150$ inhibitory units. The excitatory population has sparse internal connectivity with connection probability $p_{ee} = 0.1$. For the synapses from inhibitory to excitatory units we use $p_{ie} = 0.25$ and for the other direction $p_{ei} = 0.25$, there are no internal connections in the inhibitory population. The weights W^{ee}, W^{ie}, W^{ei} are initialized with a uniform distribution between 0 and 1, and scaled so that both the incoming excitatory and inhibitory weights of a neuron each sum to one.

All reservoir neurons are binary threshold units [van Vreeswijk & Sompolinsky 1996]. The activation y of an inhibitory neuron j at time $t + 1$ is:

$$y_j(t+1) = \begin{cases} 1, & \text{if } \sum\limits_k W_{kj}^{ei} x_k(t) > \theta_j \\ 0, & \text{else} \end{cases} \tag{5.2}$$

The thresholds θ_j are initialized in the following way: The excitatory thresholds θ^{e} are all set to a value of $\theta^{\mathrm{e}}_{\mathrm{max}}$, the thresholds of the inhibitory neurons are randomly drawn from a uniform distribution in $[\theta^{\mathrm{i}}_{\mathrm{min}}, \theta^{\mathrm{i}}_{\mathrm{max}}]$, those of the output are set to θ^{out}. The excitatory activity x uses the current inhibitory

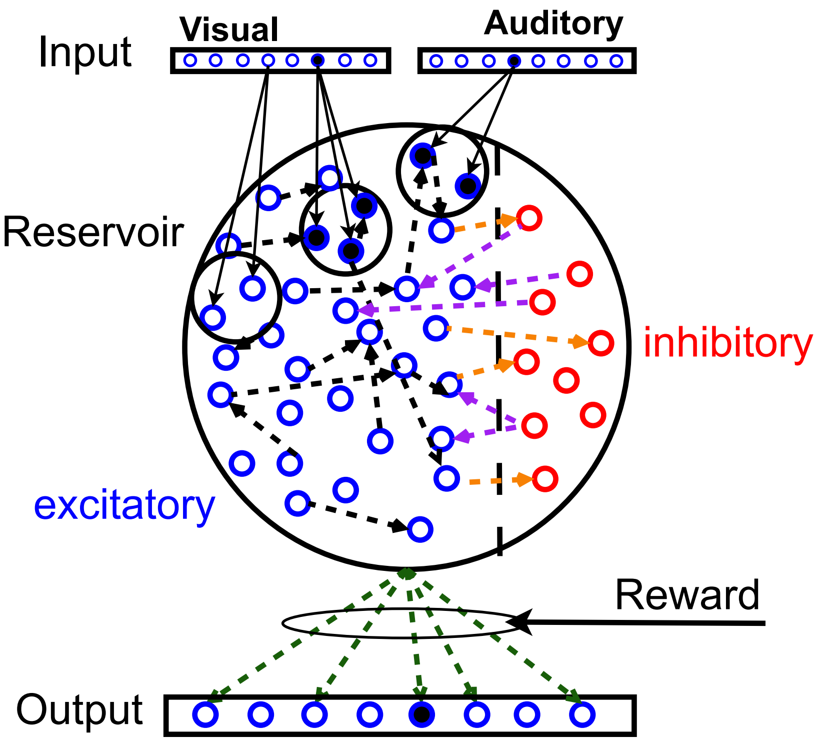

Figure 5.3. Diagram of the reservoir network. Based on the true position(s), noisy visual and auditory inputs are drawn. Among the excitatory reservoir population (blue) are small groups of neurons each coding for a single position in one modality (small circles inside the reservoir). The two groups representing the current input will have an activity of 1. All neurons in the reservoir are randomly and sparsely connected with other excitatory neurons, as well as with the pool of inhibitory neurons (red). All connections are directed. There is one output unit per position and these are connected with all excitatory units. All neurons are binary-threshold units. Among the output neurons that reach the threshold, there is a winner-take-all mechanism, that is only the unit with highest potential is allowed to spike.

signals and the excitatory signals from the last time step:

$$x_j(t+1) = \begin{cases} 1, & \text{if } \sum_{k_1} W^{ee}_{k_1 j} x_{k_1}(t) - \sum_{k_2} W^{ie}_{k_2 j} y_{k_2}(t+1) + U_j(t+1) > \theta_j \\ 0, & \text{else} \end{cases} \tag{5.3}$$

$U_j(t+1)$ is the current external input. Each possible input position is represented by one group of excitatory reservoir neurons of size ν_{inp} for each modality, which receive an external stimulation of size 1 if a stimulus shows up at their position. Additionally, $U(t+1)$ also includes a background noise term that is drawn for each neuron from a uniform distribution on $[0, \phi]$.

The output population has $|Z|$ units, one for each object position. It as well consists of binary threshold units, which receive inputs from all excitatory neurons in the reservoir. The weights W^{out} are also initially drawn from a uniform distribution in $[0, 1]$ and normalized so that the incoming weights of each neuron sum to one. Each output neuron receives $N_e + 1$ inputs, the additional weight connects it with the go signal for the decision which is 1 if $t = dT + 1$ and 0 otherwise. The outputs z'_j use the excitatory activation from the current timestep to compute their potential.

$$z_j'^{\text{pot}}(t+1) = \sum_k W_{kj}^{\text{out}} x_k(t+1) \tag{5.4}$$

Among those units with potentials higher than their thresholds we use a WTA mechanism to determine the one winning unit that is allowed to spike.

$$z_j'(t+1) = \begin{cases} 1, \text{ if } z_j'^{\text{pot}}(t+1) > \theta_j \text{ AND } z_j'^{\text{pot}}(t+1) = \max\limits_l(z_l'^{\text{pot}}(t+1)) \\ 0, \text{ else} \end{cases} \tag{5.5}$$

Based on the distance between the winning unit and the true position a reward is computed according to eq. (5.1). Each trial takes T timesteps, before a new stimulus is drawn. We do not reset the network activity between trials, that is there is potential spill-over from previous trials.

5.2.3 Plasticity Mechanisms

For the training of the output weights W^{out} we use an R-STDP rule similar to [Izhikevich 2007]:

$$\delta W_{ij}^{\text{out}}(t+1) = \eta^{\text{out}} \delta r(t+1)(\sum_{t'=0}^{\infty} e^{\frac{-t'}{\tau_e}} z_j'(t-t'+1)x_i(t-t') - z_j'(t-t')x_i(t-t'+1)) \;\forall\; i \in [1, N_e] \tag{5.6}$$

In the above equation and in the following we will use η for the relevant learning rate, the values for these and all other constants can be found in table 5.1. The sum holds the exponentially decaying eligibility trace from the original rule. Note that the prediction error $\delta r(t)$ (and by that $\delta W_{ij}^{\text{out}}(t)$) can only be different from zero when an actual reward is given (which happens only if $dT + 1 \equiv (t \bmod dTrial) - 1$). For our output weight learning we set $\tau_e \approx 0$ because output activity before the end of the delay dT does not have an influence on the reward and has therefore not to be reinforced. Instead of the explicit reward signal, we use the reward prediction error, which more correctly resembles both theoretical aspects of RL (see Section 2.3.1) and the signal that is thought to be carried by the dopamine release, modulating plasticity in the brain (see Section 2.3.2). We use an actor-critic architecture, where the recurrent network plays the role of the actor deciding based on the current input which action to take. To reduce the complexity the critic will compute a predicted reward given the chosen action using a simple temporal average of the last t_R steps with the given action. In contrary to the standard formulation of a critic, we do not compute any input or state dependence for its predictions. The results of our network with such a pure action-critic are much better compared to using a state-dependent critic or one that just averages over all trials (data not shown).

For the weights $W_{N_e+1j}^{\text{out}}$ from the go cue we use a different learning rule that is inspired by homeostatic mechanisms and tries to make all output units fire with a similar frequency.

$$\delta W_{N_e+1j}^{\text{out}} = \eta^{\theta^{\text{out}}}(z_j'(t+1) - \frac{1}{|Z|}) \tag{5.7}$$

The excitatory units in the reservoir use a variety of plasticity mechanisms. The weights W^{ee} are adapted using a simple discrete STDP rule:

$$\delta W_{ij}^{\text{ee}} = \eta^{\text{ee}}(x_j(t+1)x_i(t) - x_j(t)x_i(t+1)) \tag{5.8}$$

Additionally, after each update these weights are normalized to make all incoming weights sum to one again ("synaptic scaling").

$$W_{ij}^{\text{ee}} = \frac{W_{ij}^{\text{ee}}}{\sum\limits_{i'} W_{i'j}^{\text{ee}}} \tag{5.9}$$

As a third mechanism the thresholds θ^{e} are changed using IP [Triesch 2005a], adapted to binary threshold units [Lazar *et al.* 2007]:

$$\theta_i^{\mathrm{e}} = \theta_i^{\mathrm{e}} + \eta^{\theta^{\mathrm{e}}}(x_i(t) - \lambda^{\mathrm{e}}) \tag{5.10}$$

We use $\lambda^{\mathrm{e}} = \frac{2\nu_{\mathrm{inp}}}{N_e}$ with the idea that at each timestep the expected number of active neurons should be of the size of the input activation. Increasing or decreasing this value by a factor of 10 degrades the network performance significantly.

Finally the connections from inhibitory to excitatory neurons are also plastic and use another form of Hebbian learning similar to what was introduced in [Sprekeler *et al.* 2011, Vogels *et al.* 2011]:

$$\delta W_{ij}^{\mathrm{ie}} = \eta^{\mathrm{ie}}(x(t)y(t+1) + x(t+1)y(t+1) - \alpha y(t+1)) \tag{5.11}$$

Weights from inhibitory neurons that are good in predicting the activity of the post-synaptic excitatory neuron will be strengthened. The intuition behind this rule is that only unpredicted activities carry relevant information and should therefore be allowed. Additionally, strengthening weights of neurons that spike in parallel (therefore not causal to each other) will result in excitatory neurons that only rarely fire in two consecutive timesteps. Inhibitory spikes that will not contribute to these goals will depress the synapse. If we only use the causal part of the formula results were significantly worse (data not shown).

Table 5.1. Default parameter values for the reservoir

Symbol	Value	Symbol	Value	Symbol	Value	Symbol	Value
t_R	100	$\lvert Z\rvert$	20	p_{ee}	0.1	λ^{e}	$0.005 + \frac{2\lvert Z\rvert}{N_e}$
$\theta^{\mathrm{e}}_{\mathrm{max}}$	0.1	T	5	p_{ei}	0.25	α	0.2
$\theta^{\mathrm{i}}_{\mathrm{min}}$	0.01	N_e	600	p_{ie}	0.25		
$\theta^{\mathrm{i}}_{\mathrm{max}}$	0.045	N_i	150	η^{out}	10^{-4}		
θ^{out}	0.01	ρ	4	$\eta^{\theta^{\mathrm{out}}}$	10^{-4}		
ϕ	0.1	η^{ie}	$5 * 10^{-5}$	η^{ee}	$5 * 10^{-5}$		

The definitions of the symbols can be found in the text. These values are used in all experiments except when stated differently.

5.2.4 Bayesian Observer Models

The Bayesian observer models for the delayed response task are basically the same as described in Section 3.2.3. The two differences are the use of a slightly changed reward function (eq. 5.1) and dt_{av} as an additional cue that can be used for inferring the number of objects in the scene. The probability of a common cause used here is therefore:

$$p(C=1|z^a, z^v, dt_{av}) = \frac{p(dt_{av}|C=1)p(C=1)\int p(z^a|x)p(z^v|x)p(x)\,dx}{p(z^a, z^v, dt_{av})} \tag{5.12}$$

There is no additional term incorporating the delay to the response dT, because we assume perfect memory.

For the accumulation task we use two types of observers, one that uses only information from the last timestep before the action, and a second that accumulates information over all dT steps. The utility functions U are always only computed at the time of the decision, since the reward function is only defined on concrete actions. An accumulation will only happen in the posteriors that are used to finally determine the utility. The unisensory, single time step observer uses:

$$U(z'|z^a) = \int r(z'|z_a^*)\frac{p(z^a|z_a^*)p(z_a^*)}{p(z^a)}\, dz_a^* \ , \tag{5.13}$$

and likewise for z^v.

Accumulating evidence in a single modality in contrast uses all observations z_t^a between timesteps $t = 1$ and $t = dT$ can be written as:

$$U(z'|z^a_{t=dT}, z^a_{t=dT-1}, \cdots, z^a_{t=1}) = \int r(z'|z_a^*)\frac{p(z_a^*)\prod\limits_{t=1}^{dT} p(z_t^a|z_a^*)}{p(z^a_{t=dT}, z^a_{t=dT-1}, \cdots, z^a_{t=1})}\, dz_a^* \ , \tag{5.14}$$

Along the same lines are the equations for the multisensory observers, for the single step version please see eq. (3.7) in Section 3.2.3 and eq. (5.12) above. In the multi-step version the observer accumulates three things in parallel, the posterior given a single cause, the posterior given multiple causes and the the probabilities of these models.

$$U(z'|z^a_{dT}, z^v_{dT}...z^a_1, z^v_1) = \begin{array}{l} p(C=1|z^a_{dT}, z^v_{dT}...z^a_1, z^v_1) \int r(z'|z^*)p(z^*|z^a_{dT}, z^v_{dT}...z^a_1, z^v_1)dz^* \ + \\ p(C=2|z^a_{dT}, z^v_{dT}...z^a_1, z^v_1) \iint r(z'|z^*_a, z^*_v)p(z^*_a|z^a_{t=dT}...z^a_{t=1})p(z^*_v|z^v_{t=dT}...z^v_{t=1})dz^*_a z^*_v \end{array} \tag{5.15}$$

The three parts can be transformed to

$$p(C=1|z^a_{dT}, z^v_{dT}, ..., z^a_1, z^v_1) = \frac{p(C=1)\int p(z^*)\prod\limits_{t=1}^{dT} p(z_t^a|z^*)p(z_t^v|z^*)dz^*}{p(z^a_{dT}, z^v_{dT}, ..., z^a_1, z^v_1)} \ , \tag{5.16}$$

$$p(z^*|z^a_{dT}, z^v_{dT}...z^a_1, z^v_1, C=1) = \frac{p(z^*)\prod\limits_{t=1}^{dT} p(z_t^a|z^*)p(z_t^v|z^*)}{p(z^a_{dT}, z^v_{dT}, ..., z^a_1, z^v_1|C=1)} \ , \tag{5.17}$$

$$p(z^*_a|z^a_{t=dT}...z^a_{t=1}, C=2) = \frac{p(z_a^*)\prod\limits_{t=1}^{dT} p(z_t^a|z_a^*)}{p(z^a_{dT}, ..., z^a_1|C=2)} \ , \tag{5.18}$$

and $p(z^*_v|z^v_{t=dT}...z^v_{t=1}, C=2)$ defined equivalently to the latter, and again $p(C=2) = 1 - p(C=1)$.

5.2.5 Supervised Read-out

To access the quality of our online reward-based output learning, we have to separate its contribution from that of the reservoir plasticities. This can be done be comparing the performance of our output with results from an offline supervised trained read-out. Such a read-out is indeed used in most of the work on LSMs/ESNs, including the original proposals [Jaeger 2001, Lukoševičius & Jaeger 2009, Maass *et al.* 2002]. We use the direct method of computing the Moore-Penrose pseudo-inverse X^+ of the matrix X of all excitatory reservoir activity vectors at time dT from 60,000 randomly generated trials in

$$W^{\text{out}} = Y_{\text{true}}X^+ \ , \tag{5.19}$$

where Y_{true} is the binary matrix holding the true input position of all the trials. The performance evaluation of this read-out on the same data as the reward-dependent output is done based on the position estimate from a WTA over the output activity. This measure sets a good upper limit for the possible performance with the trained reservoir in terms of the least squared error.

5.3 Results

5.3.1 Cue Integration & Causal Inference

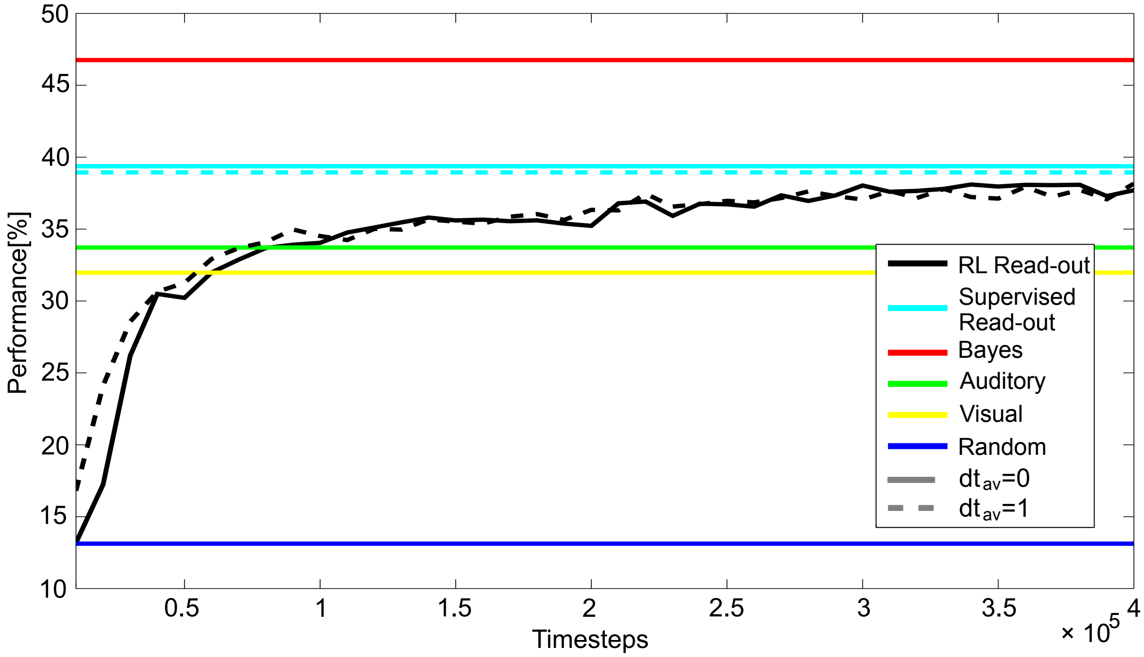

Figure 5.4. Performance of the network during training. Plot of the performance of the reservoir network with all plasticity mechanisms and R-STDP at the output weights (black) with optimal Bayesian integration of the two cues (red), the two unisensory observers (green and yellow), performance at chance ("random", blue) and the same reservoir with supervised output learning (cyan) ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$). Solid black line: Task with $dT = 1$, $dt_{av} = 0$. Dashed black line: Task with $dT = 1$, $dt_{av} = 1$.

We start with describing the results on the **delayed response task**. As Fig. 5.4 shows, the network is able to successfully learn to integrate two synchronous (no intercue-delay, $dt_{av} = 0$) cues for $dT = 1$. The performance is larger than that predicted by a pure unisensory observer, which means that information from both cues is used by the network. Unfortunately the Bayesian observer can get more reward than the network, but note that we did not include any delay penalty into the optimal model.

To disentangle the contributions of output versus reservoir learning, we can also compare our results with those of a read-out that is trained offline in a supervised manner. Most previous work on reservoir networks used such methods [Jaeger 2001, Lukoševičius & Jaeger 2009, Maass *et al.* 2002], and it is therefore a good comparison to evaluate the more biological reward-modulated plasticity we are using. For more details on the supervised read-out training please see Section 5.2.5. As can be seen in Fig. 5.4, the performance of both methods is similar.

Due to physical delays and differences in processing time natural stimuli from different modalities will often arrive asynchronously in an integration area in the brain. Humans are able to make use of both signals, even if the delay is bigger than the integration time constant of a neuron (e.g. [Child & Wendt 1938]). We simulate such a setup in our memoryless neurons by imposing an offset between the two cues, e.g. auditory after visual ($dt_{av} = 1$). The performance does not change much (Fig. 5.4 dashed line).

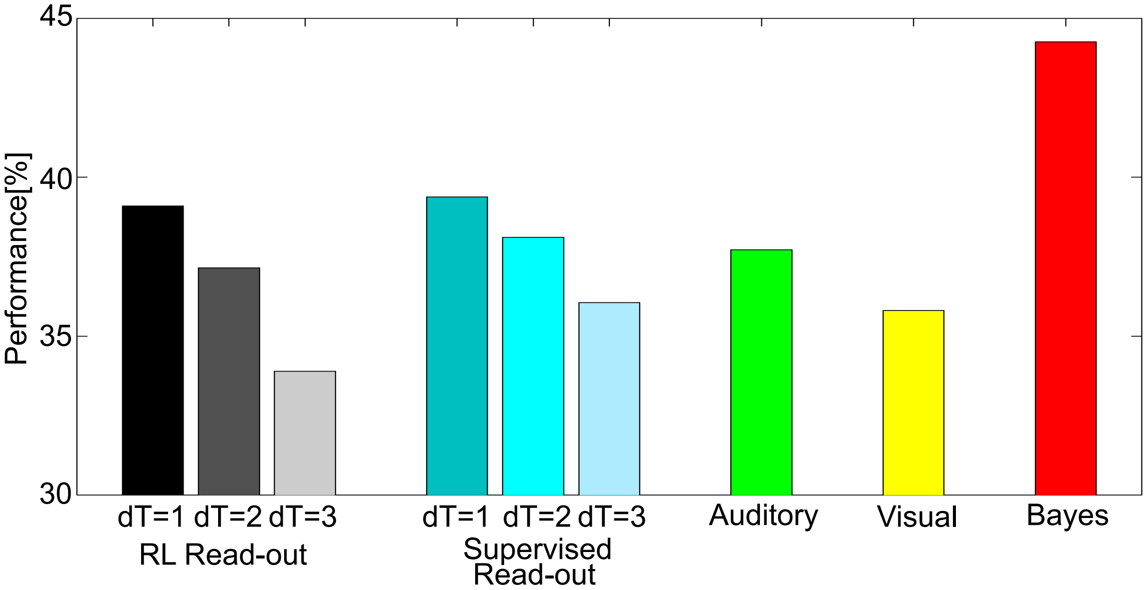

Figure 5.5. Performance for different delays. The plot compares the performance of both a reward modulated (gray) and a supervised offline trained (cyan) read-out with trained reservoirs for delays $dT = 1, 2, 3$. The performance of both read-outs declines with increasing delay, though the difference between the read-outs seems to also grow. We do not show error bars in this and other plots because of long simulation times. Nevertheless, for those cases that were run for a bigger number of times we never found inter-run variations bigger than ±0.5%, which is of the order of the intra-run variations after convergence. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $dt_{av} = 0$).

Figure 5.5 shows the performance of the network for different delay times dT. With increasing waiting time the performance of both the reward-dependent and the supervised read-out decline. This effect demonstrates a limitation in the memory of the reservoir, something already seen in previous work [Lazar *et al.* 2007]. Additionally, we can also see that the difference between supervised and reward-modulated output learning increases with increasing delay.

As we have done in the previous Chapters, we also run the network with a setup including causal inference, where the stimuli either arise from a common or two independent causes ($p(C = 1) = 0.5$). The performance in this task is similar to the simpler version, again the network is slightly better than the single cues (at least when looking at the supervised read-out), but not as good as the Bayesian predictions (Fig. 5.6). We plot the performance for $dT = 2$, where as we have seen before the difference between reward-mediated and supervised read-out is clearly visible, the plot for $dT = 1$ is very similar to Fig. 5.4. Figure 5.6 also shows the results for delays between the cues (dashed lines), but here this delay could also be used as another cue for causal inference (see eq (5.12)). If a trial has a single audio-visual stimulus, we set $dt_{av} = 0$, for the case with two causes dt_{av} is drawn from a uniform distribution over $\{-1, 0, 1\}$. As we can see, including timing as a signal seems to not have a big influence on the performance of the Bayesian observer or the network.

Figure 5.7 shows the similarity between the behaviour of the network and the Bayesian observer, and the distribution of errors that they both make. Most of the times the network decides similar to the optimal predictions, but clearly is more noisy. If we compare the similarity of the behaviour of our model with the two alternative Bayesian observers for causal inference, model averaging (MA) and model selection (MS), we do not see any systematic differences (data not shown). This is most probably due to

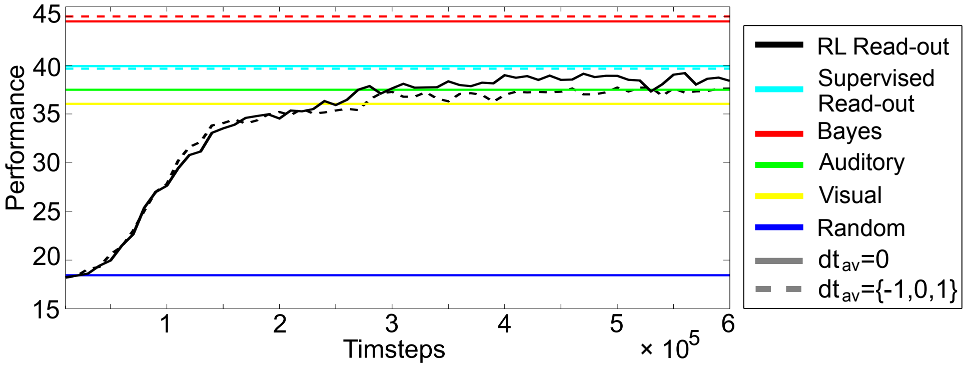

Figure 5.6. Performance of the network during training with causal inference. Plot of the performance of the reservoir network with all plasticity mechanisms and R-STDP (black) or supervised learning (cyan) at the output weights. Solid lines: Task with $dt_{av} = 0$. Dashed lines: Task with $dt_{av} = 0|C = 1$,$dt_{av} = \mathcal{U}(\{-1, 0, 1\})|C = 2$. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$,$dT = 2$).

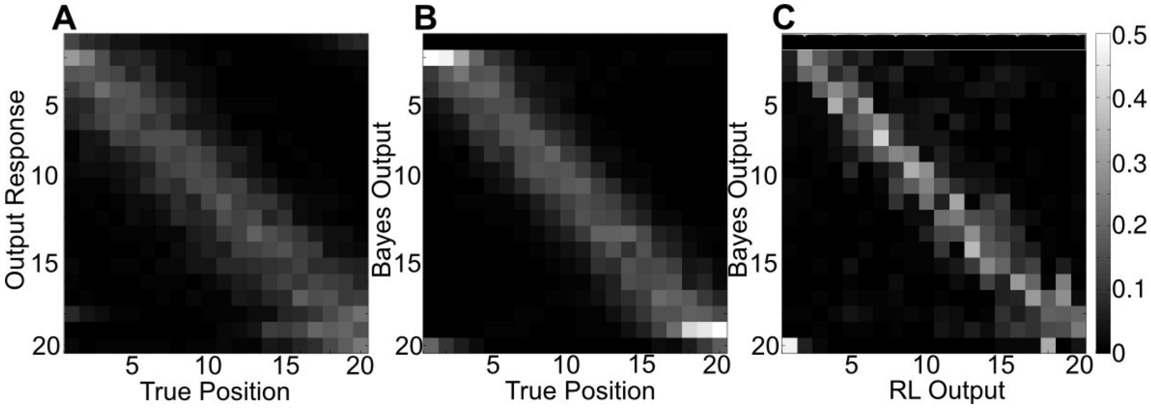

Figure 5.7. Confusion matrices. A: Comparison of the output of the RL read-out with the true input positions. **B**: Comparison of the output of the Bayesian observer with the true positions. **C**: Comparison of the output of the RL read-out with the output of the Bayesian observer. Plots are normalized so that all rows sum to one. Brighter Colour means higher probability. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$,$dT = 2$,$P(C = 1) = 0.5$).

the larger difference in performance from both Bayesian predictions, especially since we have already seen in Chapter 3 that difference between these two versions for the given task are only visible from detailed analysis.

To be able to assess the contributions to encoding of the stimuli, we compute the "depth of selectivity" (DoS) of each neuron, a measure that is often used in experimental work [Rainer *et al.* 1998]. This measure compares the firing probability of the stimulus $s \in S$ that drives a neuron most with the overall firing probability of this neuron:

$$\mathrm{DoS} = \frac{|S| - \frac{\sum_{i=1}^{|S|} <x>_{(t|s_t=i)}}{\max_i(<x>_{(t|s_t=i)})}}{|S| - 1}, \tag{5.20}$$

where $< x >_{(t|s_t=i)}$ denotes the mean activity of a neuron over all timesteps in which the stimulus was i. A DoS of 0 would mean that the mean activity is the same for every stimulus, a 1 that the neuron

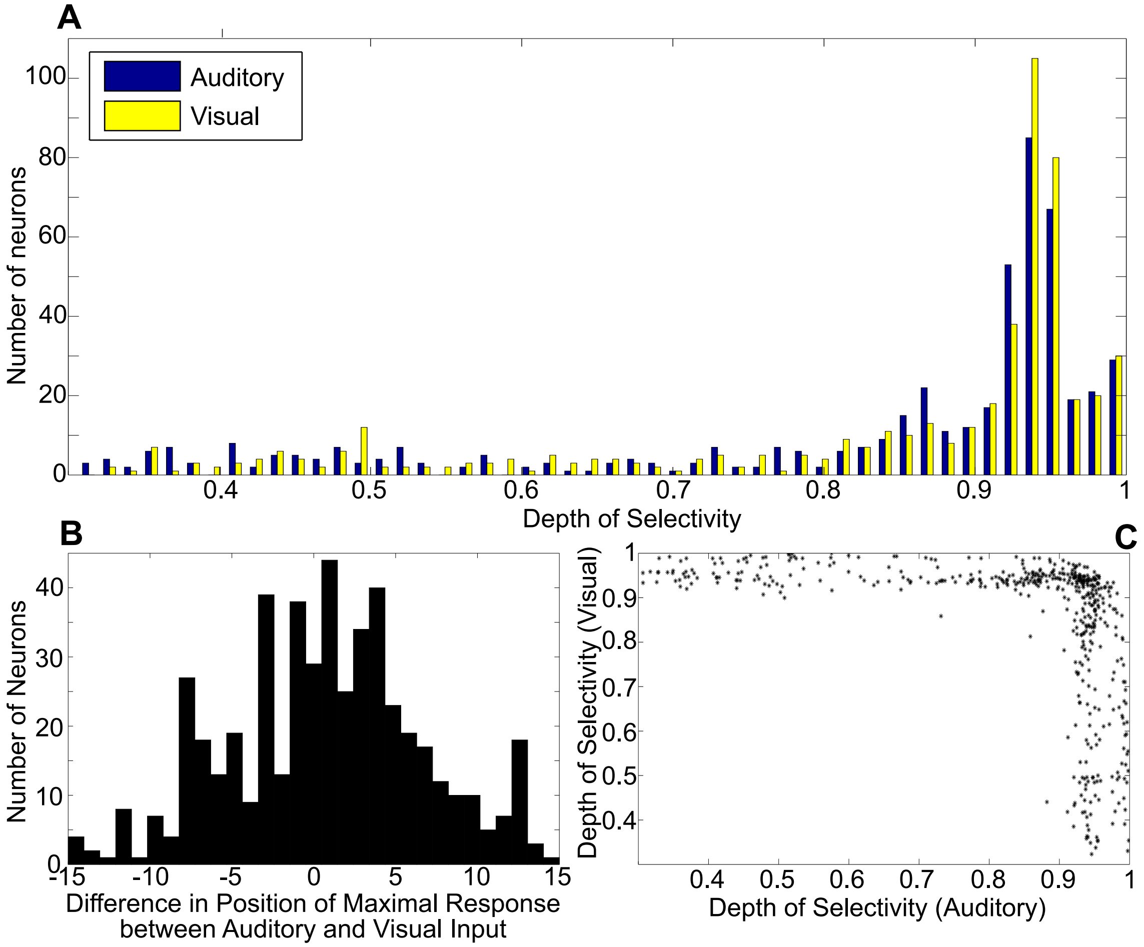

Figure 5.8. Selectivity of reservoir neurons to input position. A: Histogram of neurons $\mathrm{DoS}_{\mathrm{inp}}$ separately for auditory and visual inputs. **B**: Histogram of the difference between the preferred auditory and visual position within the same neuron. **C**: Correlation between auditory and visual $\mathrm{DoS}_{\mathrm{inp}}$. Each point marks one reservoir neuron. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$,$P(C = 1) = 0.5$).

is only responding to a single stimulus at all. It is possible to define selectivity both based on the input stimuli or on the true hidden states. The depth of selectivity for input stimuli ($\mathrm{DoS}_{\mathrm{inp}}$), both auditory and visual, is very high for many neurons (Fig. 5.8A). Most neurons actually react almost exclusively to a single position in a given modality. We can compare $\mathrm{DoS}_{\mathrm{inp}}$ and the position of maximal response between the auditory and visual input. The biggest group of neurons is equally selective for one position in both modalities (Fig. 5.8C), but there are also a number of neurons more specific for one of the cues. Figure 5.8B shows the similarity in the position of maximum tuning of the same neuron for auditory and visual position. The distribution shows a clear peak around zero and a preference for small differences which could reflect features of the generating model for a common cause.

Alternatively, we check if the neurons also exhibit specificity for the hidden variables. The most relevant variable for the task used here is the true position of the stimulus ($\mathrm{DoS}_{\mathrm{pos}}$), since it is the value which the read-out has to decode from the network's activity. Figure 5.9 shows the DoS for this variable for all reservoir neurons, ordered by the true position that evokes highest activity. For every position one can find a number of neurons with high specificity for it. The number of neurons for very low and high

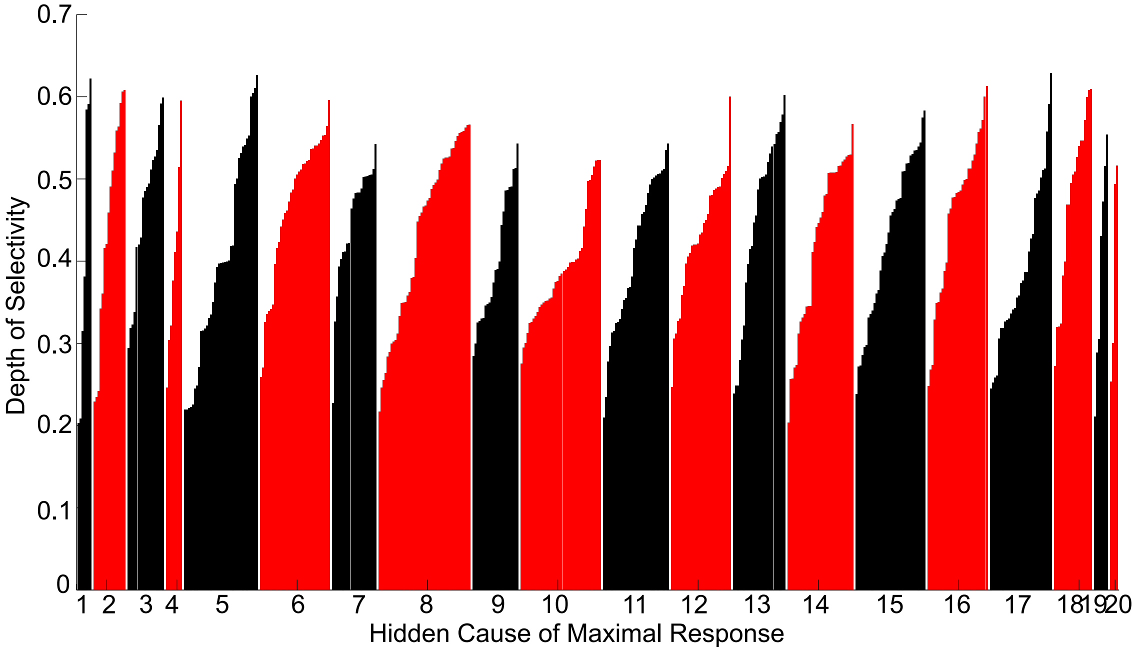

Figure 5.9. Selectivity of reservoir neurons to true stimulus position. Each bar shows the depth of selectivity of a single reservoir neuron. The neurons are first ordered by the stimulus of maximum activity (colors alternate to enhance visibility of groups with same max-stimulus) and then in ascending order of DoS. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$,$P(C = 1) = 0.5$).

positions is smaller due to the border effects, which will often lead the noise to push the input stimulus out of the range, which results in the network not receiving an input.

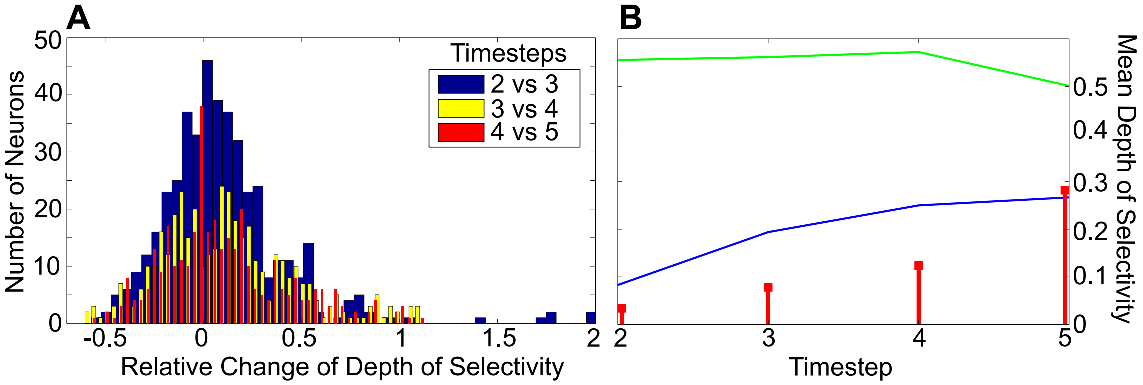

Figure 5.10. DoS_{pos} **changes over time.** The plots show how the depth of selectivity for the true stimulus position evolves over the 5 timesteps of a trial. **A**: Histogram of the increases and decreases of DoS_{pos} between consecutive timesteps, relative to the value in the first step. **B**: The average DoS_{pos} over all active neurons at each step (green) stays relatively constant. The average DoS_C increases with time (blue). The red bars show the fraction of neurons that are not responsive at all in the respective timestep. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$,$P(C = 1) = 0.5$).

Since our neurons do not have any memory, each neuron could actually code for a different stimulus at different time points. It is also possible, that selectivity grows with time both due to selective inhibition or combining signals from multiple already specific neurons. Figure 5.10A shows the distribution of relative changes in $\mathrm{DoS}_{\mathrm{pos}}$ between two consecutive timesteps within a trial (relative to the value in the previous step). The means of the distributions are +9.6% between second and third timestep, +12.4% between third and fourth and +12.3% between fourth and fifth timestep. The average $\mathrm{DoS}_{\mathrm{pos}}$ of a neuron for each timestep is shown in Fig 5.10B (green curve). Note that from the differences between these two plots we can deduce that mostly neurons with small values increase whereas those with higher values decrease in $\mathrm{DoS}_{\mathrm{pos}}$. In absolute values these changes cancel out each other (therefore the almost constant curve in Fig. 5.10B), but the above explains the positive bias in relative values. For both plots we removed those neurons that never fired in that particular (pair of) timesteps. The number of those neurons increases with time (Fig 5.10B, red bars).

In the causal inference setup, there exists an additional hidden variable (C). The blue curve in Fig 5.10B shows the average selectivity of the neurons for this value DoS_C in each timestep. Despite being relatively small it shows a clear increase with time.

5.3.2 Effect of the Plasticity Mechanisms

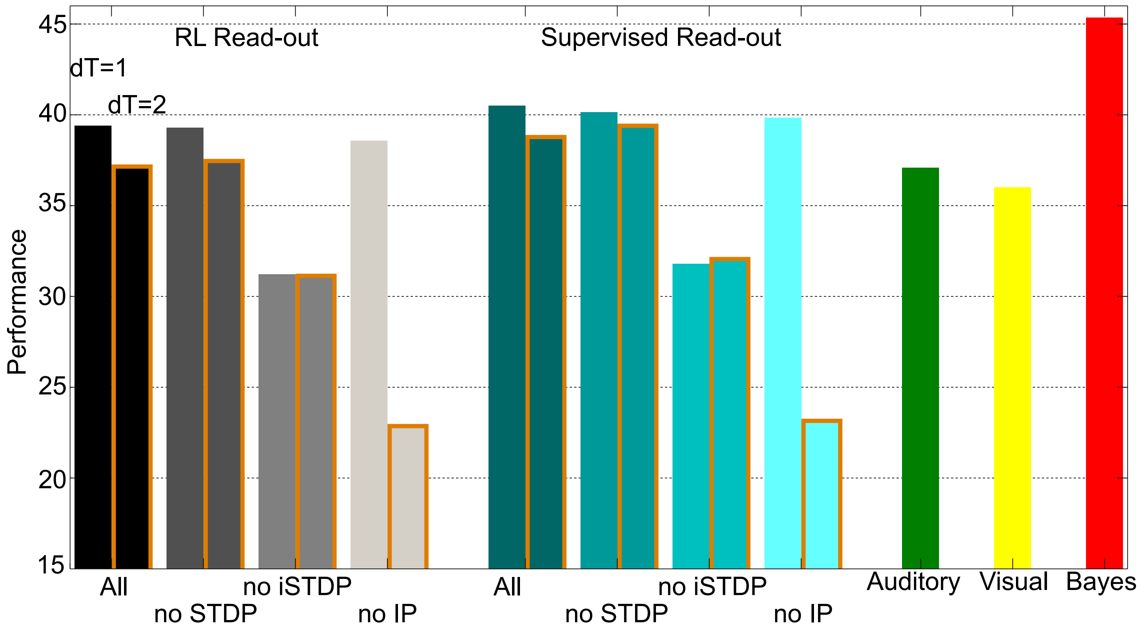

Figure 5.11. Performance of the recurrent network with different combinations of plasticity mechanisms. The plot compares the performance of both a reward modulated ("RL") and a supervised offline trained read-out with different reservoirs for delays $dT = 1$ and $dT = 2$ (bars with orange edges). The reservoirs are either simulated with all plasticity mechanisms active ("All") or with each one of them turned off separately. As is shown in Fig. 5.5, performance decreases for the longer delay in most cases. Without inhibitory plasticity (iSTDP) performance decreases for both delays, the positive effect of IP only shows for the longer one. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$).

We were interested in the contributions of the individual plasticity mechanisms to the performance of the network. Figure 5.11 shows the performance of four reservoirs, each using a different subset of plasticity mechanisms. Interestingly, it seems that STDP does not improve (nor worsen) the reservoir

(Fig. 5.11), which is different to previous results on similar types of networks [Lazar *et al.* 2007, Lazar *et al.* 2009].

In contrast, if we run the network without inhibitory plasticity (that is only IP (data not shown), or IP and STDP) performance converges to a much smaller value. Nevertheless the output weights are still able to capture some regularities in the reservoir's state. Turning off IP does not have an influence on the performance for a delay of 1, however when increasing dT its effect is quite drastic (Fig. 5.11).

The effect of the different plasticity mechanisms can also be seen from the activity patterns of the reservoir (Fig. 5.12). Networks that use both IP and iSTDP (Fig. 5.12 A and B) show a distribution of the activity over multiple timesteps after stimulus input ($t = 1$). The effect of STDP on those networks seems to strengthen short delay activity at the cost of later timesteps. The reason for this effect can be found in the type of changes on the excitatory weights (Fig. 5.13), we will get back to that in some more detail later in this section. Turning off IP destroys the diversity of firing behaviour among neurons, e.g. they all fire in the first timestep after the input and therefore can not carry on information to later steps (Fig. 5.12C). Note that the exact mean activity pattern is depending on the initial value for the thresholds of the excitatory reservoir units (in Fig. 5.12C we use $\theta^{\mathrm{e}}_{\max} = 0.1$ as in all other simulations). We tested the network with a variety of thresholds without improving the performance. The mean activity patterns found for an alternative setting (θ^{e} normal distributed around a "good" value) can be seen in appendix A, Fig. A.1. For very high or low θ^{e} we find no activity or constant firing respectively (data not shown). A network without iSTDP shows oscillations between high and low population activity. After the strong input stimulation it seems as if the inhibitory population shuts down most excitatory neurons in the next step. This silence in the second step of a trial explains the low performance of this network

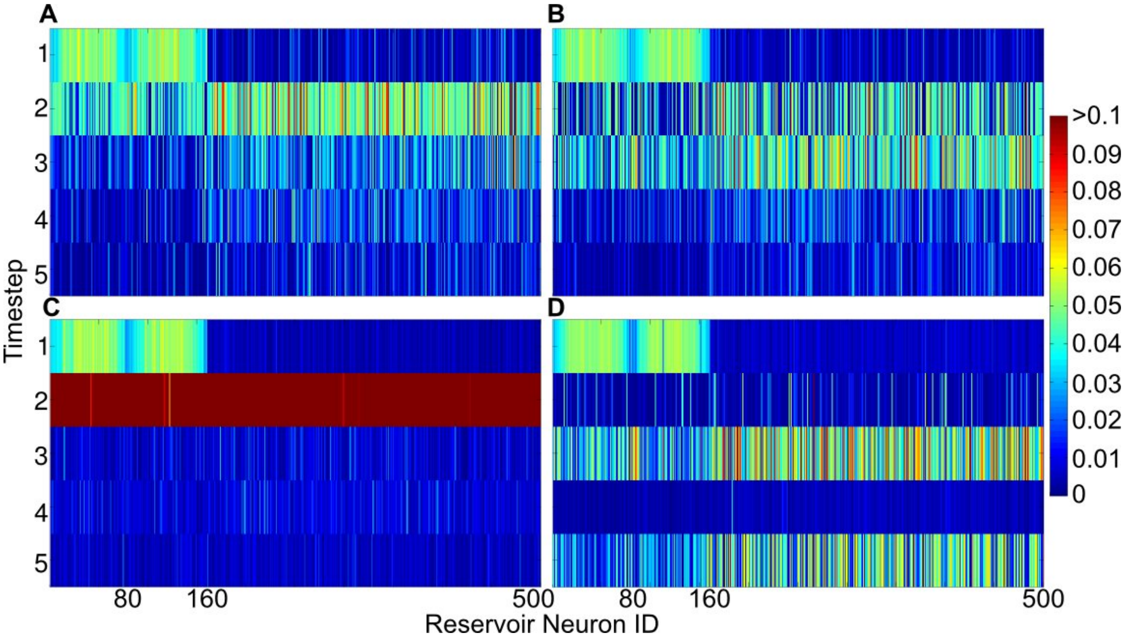

Figure 5.12. Average firing of the reservoir neurons over all inputs with different combinations of plasticity mechanisms. The figure shows the firing of each excitatory neuron in the reservoir averaged over 10,000 stimuli. Each subplot shows a different plasticity setup: **A**: All plasticities active. **B**: Only IP and iSTDP active. **C**: Only STDP and iSTDP active. **D**: Only IP and STDP active. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$, $dT = 2$).

(see Fig. 5.11), since most of the input information will get lost.

IP and iSTDP seem to address different forms of sparsity in the neuronal firing. IP was designed to lead to lifetime sparsity of each neuron separately, that is to only fire rarely. iSTDP strengthens population sparsity, by not allowing big groups of neurons to fire at the same timestep, although this seems to only work with active IP. Both of these effects were also found in the brain (e.g. [Baddeley *et al.* 1997, Olshausen & Field 2004, Perez-Orive *et al.* 2002]). Histograms of both population and lifetime activity, specifically with respect to the contributions of these to plasticity mechanisms can be found in Fig. A.2 in the appendix.

We can also take a closer look at the direct impact of the plasticity rules. Figure 5.13 for example shows the weight matrix for connections within the excitatory pool of neurons before and after training. These values are directly influenced by STDP and synaptic scaling. STDP strengthens the connections from input population neurons. Due to at the synaptic scaling mechanism this will concurrently depress all other synapses.

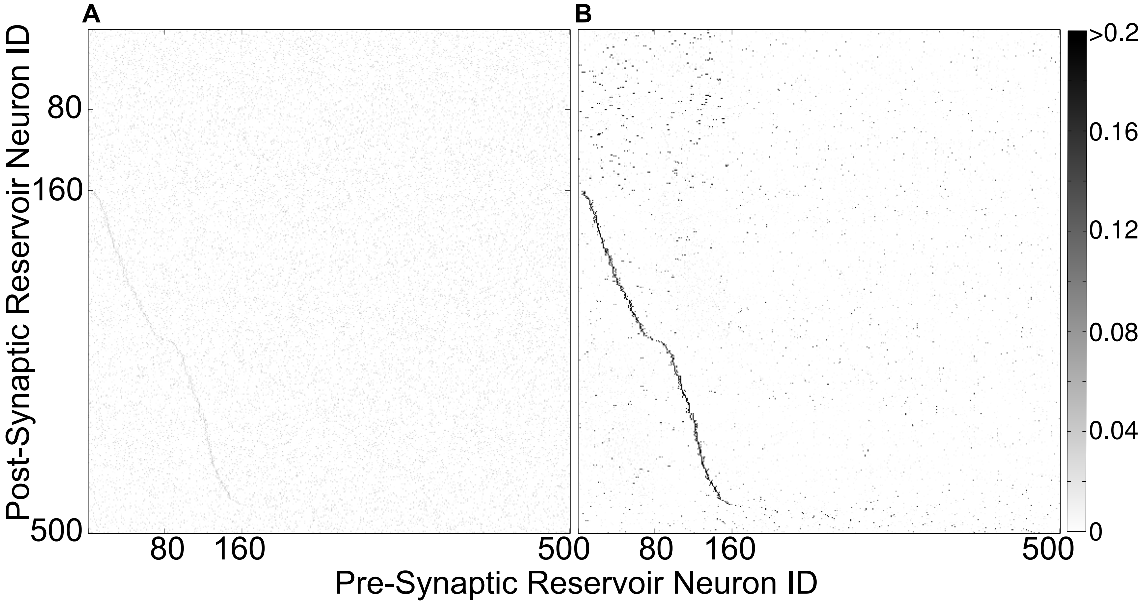

Figure 5.13. Excitatory weight matrix before and after STDP learning. Connection strength of all synapses between excitatory units (W^{ee}). Units with ID 1 to 80 are from the auditory input population, those with ID 81 to 160 from the visual input population. The rest of the neurons in both plots (rows) are sorted by the ID of the neuron with highest incoming projection weight after training. **A**: Random initial weights. **B**: After training with STDP and synaptic scaling for $500,000$ iterations (iSTDP and IP were also active). ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$)

In Fig. 5.14 we can see an even stronger effect of iSTDP on the projection weights from inhibitory to excitatory neurons. After learning, the weight matrix is dominated by few very strong synapses, basically every excitatory neuron receives non-zero synapses from only a single inhibitory unit. This result, in combination with the seemingly positive influence of iSTDP on performance, is in strong contrast to connectivity patterns used in previous work with similar networks. It is usually assumed that a high connectivity between inhibitory and excitatory units (compared to the sparsity in excitatory-excitatory connections) is supportive for the reservoir's dynamics [Lazar *et al.* 2007, Lazar *et al.* 2009, Savin 2010]

We can also find traces of the mutual influences of the plasticity rules in the synaptic weights. Figure

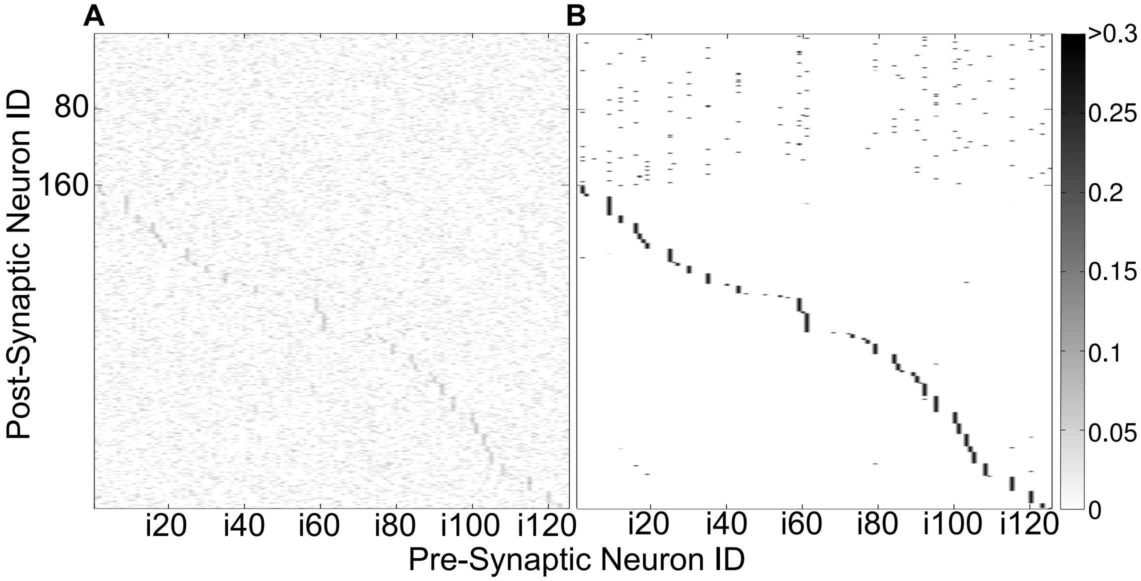

Figure 5.14. Inhibitory to excitatory weight matrix before and after iSTDP learning. Connection strength of all synapses from inhibitory to excitatory units (W^{ie}). Units with ID 1 to 80 are from the auditory input population, those with ID 81 to 160 from the visual input population. IDs with an "i" in front denote inhibitory neurons. The rest of the excitatory neurons in both plots (columns) are sorted by the ID of the neuron with highest incoming projection weight after training. **A**: Random initial weights. **B**: After training with iSTDP and synaptic scaling for 500, 000 iterations (STDP and IP were also active). ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$)

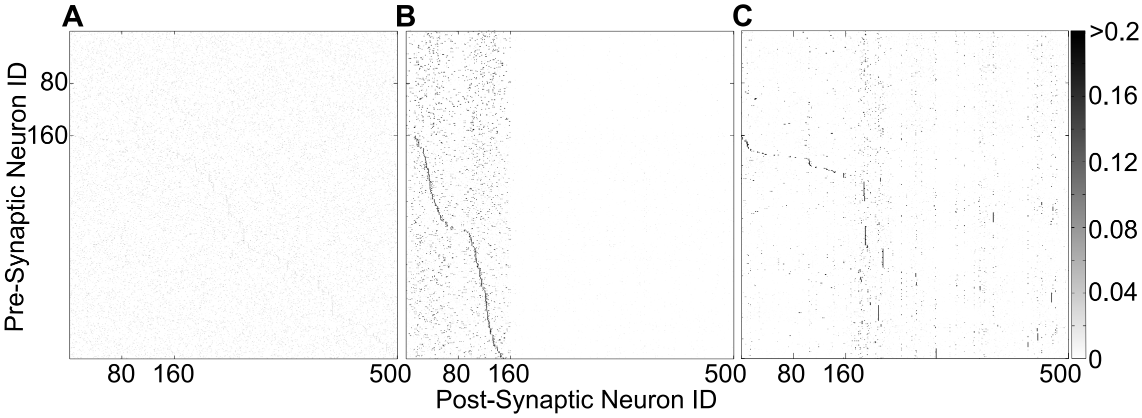

Figure 5.15. Change of the excitatory weight matrix after STDP learning depending on additional plasticities. Despite the fact that STDP and synaptic scaling are the only two plasticity mechanisms directly changing the weights of excitatory-excitatory synapses, the presence of other adaptations still has an indirect influence on those synapses. This plot shows the effects of training a network without IP or iSTDP on the final weight matrix W^{ee} (compare the results to Fig. 5.13 where all plasticities are active). **A**: Random initial weights. **B**: STDP, iSTDP and scaling but no IP. **C**: STDP, IP and scaling but no iSTDP. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$)

5.15 compares the excitatory weight matrices after STDP of networks that were trained without either IP or iSTDP. As can already be seen in Fig. 5.12, a reservoir without IP tends to only fire in the timestep right after the input. The only correlated activity will therefore be between the input population and the rest.

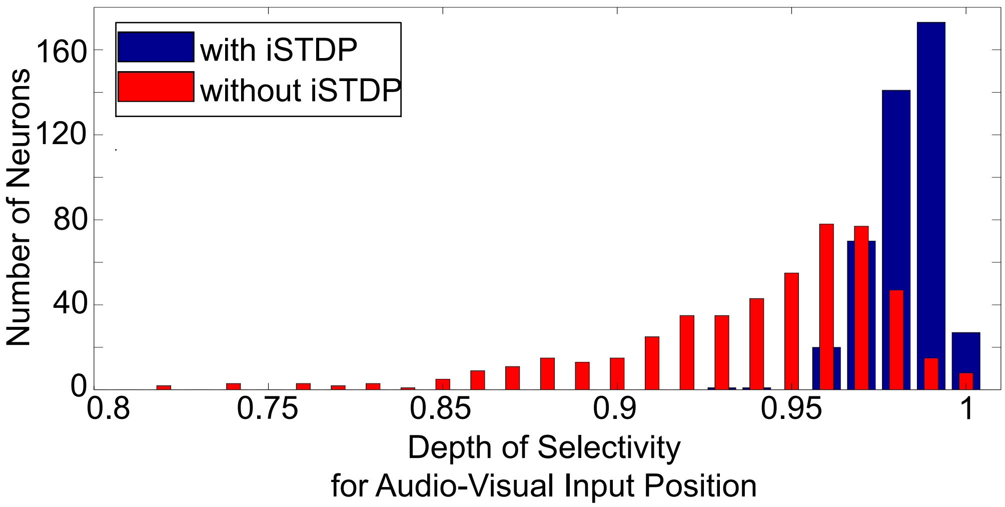

Figure 5.16. Depth of selectivity for input pairs with and without iSTDP. Histogram of the $\mathrm{DoS}_{\mathrm{inp}}^{av}$ of excitatory reservoir neurons after training with STDP, synaptic scaling, IP and with (blue) or without (red) iSTDP. Inhibitory plasticity increases the selectivity for input pairs. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$)

Bourjaily and Miller [Bourjaily & Miller 2011b, Bourjaily & Miller 2011a] found an effect of inhibitory plasticity on the selectivity of neuronal tuning. They used a slightly different mechanism, based on experimental data from [Maffei *et al.* 2006], termed LTPi, which increases inhibitory weights when the pre-synaptic cell fires and the post-synaptic neuron is depolarized but silent Additionally they also use synaptic scaling, similar to our work. Despite the differences in the learning rule and a much simpler neuron model used here, we also computed the depth of selectivity for pairs of input stimuli ($\mathrm{DoS}_{\mathrm{inp}}^{av}$) to compare the qualitative results with the results in [Bourjaily & Miller 2011b]. Figure 5.16 shows histograms of the $\mathrm{DoS}_{\mathrm{inp}}^{av}$ for one reservoir trained with all our plasticity mechanisms (STDP,IP,synaptic scaling, iSTDP) and for one were we disabled iSTDP. High values of $\mathrm{DoS}_{\mathrm{inp}}^{av}$ signal strong selectivity for exactly one combination of audio-visual input stimuli. Similar to the results in [Bourjaily & Miller 2011b] we find an increase in pair selectivity when using inhibitory plasticity. In contrast turning off STDP while iSTDP is active does not influence the distribution at all (data not shown).

In [Savin 2010] the author used R-STDP on synapses between excitatory neurons in the reservoir and showed that this improves the performance over static networks or those only using STDP. We test if this improvement could also be found in our more complex task. Alternatively to the modulation by the absolute reward value used in [Savin 2010], we test the modulation by the temporal difference (TD) error that was also used for training of the output weights (referred to as TD-STDP, see eq (5.6)). Synaptic scaling is used in all cases. Figure 5.17 shows the performance of the network after training with either one of the R-STDP rules or with standard STDP. We find no difference between R-STDP and STDP. And as was shown in Fig. 5.11, STDP itself does not improve the performance of a network without excitatory-excitatory plasticity. Interestingly, TD-STDP, which seems to work well for the output weights is not as beneficial as the alternative rules.

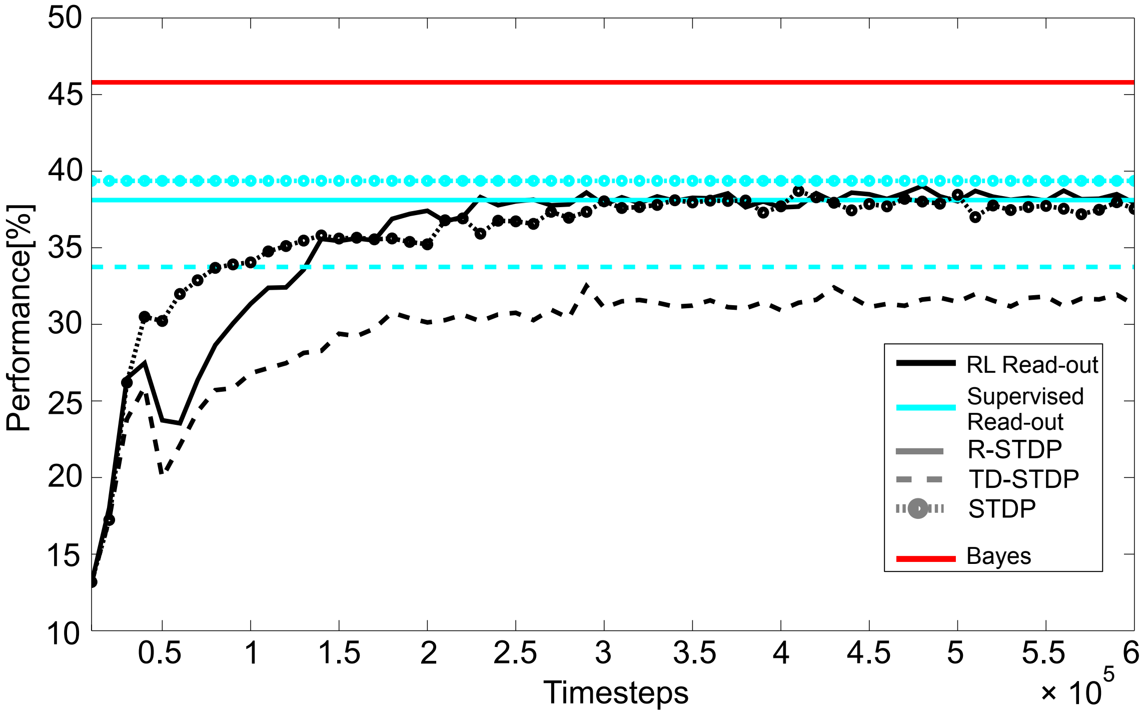

Figure 5.17. Performance of the recurrent network with different plasticity mechanisms at excitatory synapses. The plot shows the change in performance with training of both a reward modulated ("RL") and a supervised offline trained read-out on reservoirs using standard STDP (circles), Reward-modulated STDP (solid) or TD-error-modulated STDP (dashed) at excitatory-excitatory synapses. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 1.0$).

5.3.3 Temporal Integration

We will now discuss the performance of the reservoir network for the **accumulation task** in which the network receives input at every timestep. The input dynamics are therefore comparable to the task of time-series prediction that is often used in work on recurrent networks [Lazar *et al.* 2009, Lukoševičius & Jaeger 2009]. In this task performance can be improved by taking into account previous inputs from the same trial. Because we use independence of the noise between timesteps, integrating over time can use the same computations as integrating over cues. Note however that we use this setup only to limit the complexity of the Bayesian observers. Since our approach is "model-free" it could in principle also learn complex dependencies in the noise of different timesteps, as we saw in the results on realistic data with the previous model (Chapter 4).

Fig. 5.18 shows the performance of the network for smaller input populations ($\nu_{\text{inp}} = 2$) and $dT = 4$, that is the action has to be performed in the fifth timestep (the last timestep, since we use $d_{Trial} = 5$). The performance again exceeds that of the single cue observer and seems to get close to the Bayesian integration result for those models that do not integrate information over time (dashed lines). But those observers that take into account all the inputs from the previous timestep of a trial (solid lines) greatly outperform the reservoir network. If we compare the absolute reward that the network receives after training on the accumulation task with what we saw in the delayed response task, we find an improvement. This could mean that the additional information provided by the extra inputs during the "delay" is in part used by the network. In the same figure we also show the performance of a reservoir

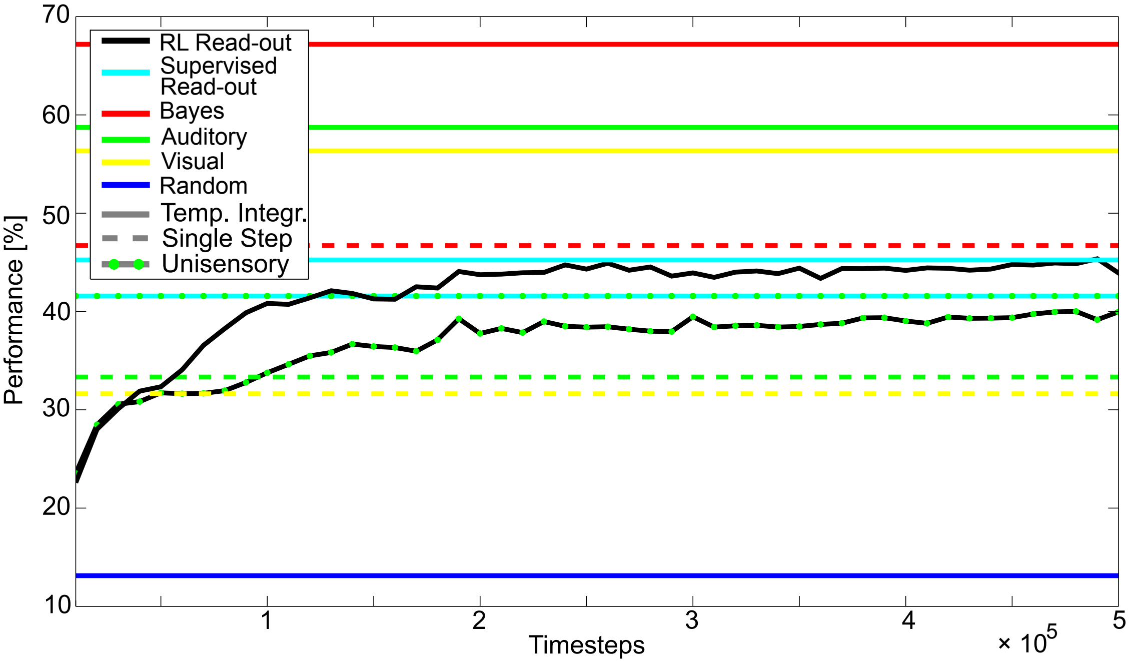

Figure 5.18. Performance of the recurrent network in the accumulation task. The plot shows the change in performance with training of both a reward modulated ("RL") and a supervised offline trained read-out in a task with noisy audio-visual input samples in every timestep. The reservoir uses STDP, iSTDP and IP and receives input at three timesteps before the output has to take an action. For comparison the performances of uni- and multisensory observers using either only information from the last timestep (dashed) or accumulated information from all three timesteps are also plotted (solid). We also plot the performance of the network for a case where only auditory signals are present ("unisensory" - lines with green dots). ($\sigma_a^2 = 3.2$, $\sigma_v^2 = 3$, $p(C = 1) = 1.0$).

that was trained only receiving auditory inputs (lines with green dots in Fig. 5.18). In this case the performance stays better than a single step auditory observer, meaning that the read-out can access information beyond what is present in the last input. But performance is also smaller than what we saw in the multisensory case. We can therefore conclude that the network integrates both spatial and temporal information. Another way to display this finding is used in Fig. 5.19. It shows the correlation between the output of the RL read-out with the actions that are taken either by the memoryless or by the temporally integrating Bayesian observer. Although there is higher overlap with the first one, for some of the action it is matched much better by the second observer. This happens mostly on the edges of the input space, where often no stimulus will be present because the noise pushes it out of the input space. In those cases it is particularly beneficial to include a second sample that might actually be present.

Figure 5.20 shows the weight after training the RL read-out in the accumulation task. First of all one clearly sees two parallel diagonal traces of strong inputs from the first 80 reservoir neurons. Those 80 neurons represent the two input populations, each of size $\nu_{\text{inp}}|Z|$. From the alignment of the neuron positions of highest input between auditory and visual input population, one can see that the read-out will integrate both cues. The weights from a second group of neurons in the center of Fig. 5.20 show a similar pattern. This is visible because we ordered the excitatory neurons (columns) by the ID of the neuron providing the highest input weight to it. Since the IDs between 1 and 80 denote the input population, we know that the input to this second group of neurons is dominated by the input population's activity

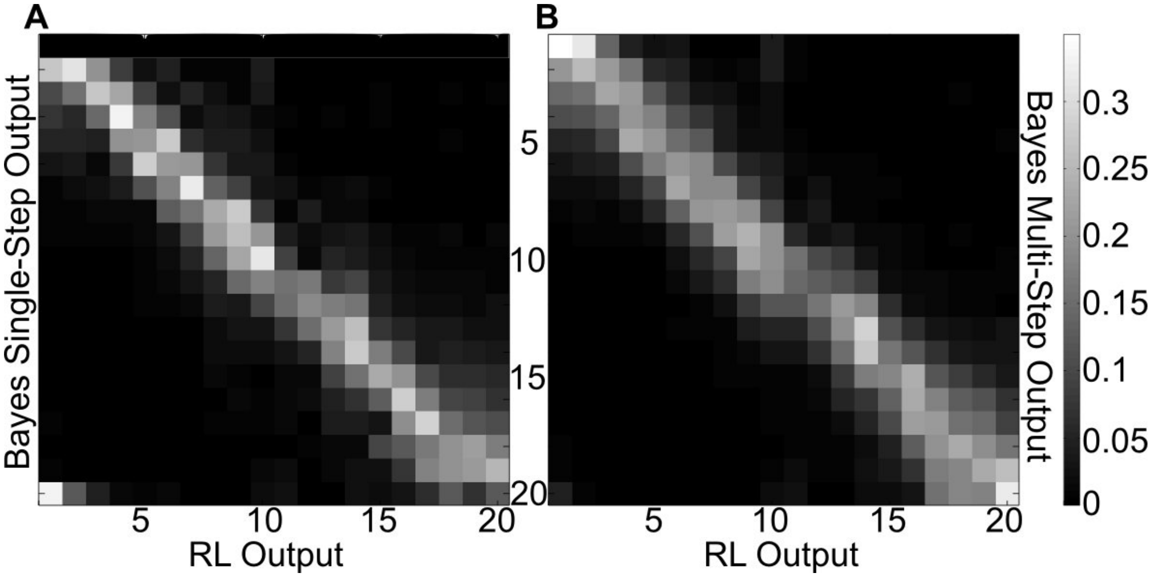

Figure 5.19. Comparison of the network with the behaviour of Bayesian observers in the accumulation task. The output of the network is compared with the output of a Bayesian observer using only information from the last timestep (**A**) and with one using accumulated information from all previous timesteps (**B**). Plots are normalized so that all columns sum to one. Brighter Colour means higher probability. ($\sigma_a^2 = 3.2$, $\sigma_v^2 = 3$, $p(C = 1) = 1.0$).

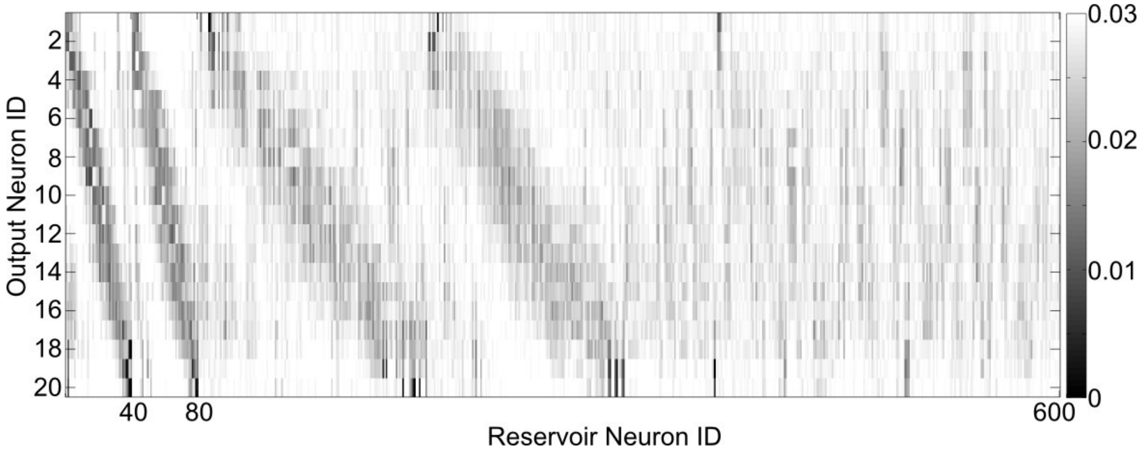

Figure 5.20. Output weights from the accumulation task. The matrix of weights from the reservoir to the output units. Units from 1 to 40 and 41 to 80 represent the input population for auditory and visual stimuli, all other columns are ordered by the reservoir neuron ID of their highest input weight. ($\sigma_a^2 = 3.2$, $\sigma_v^2 = 3$, $p(C = 1) = 1.0$).

and therefore by the stimuli one timestep earlier. This weight trace is again a diagonal, which shows a temporal integration of aligned positions over time.

The simulations in the accumulation task use the same parameters as in the delayed response task with one exception: We changed the initialization procedure for the parameter λ^{e} that is controlling the desired firing rate used by the IP rule in the reservoir. Instead of using a single target value for all neurons, we sample for each non-input unit from an exponential distribution ($exp(\frac{\nu_{\mathrm{inp}}}{N_e})+0.005)$). For the units receiving input stimulation we set $\lambda^{\mathrm{e}} = \frac{1}{|Z|}$ because we want them to consistently spike in response to a stimulus. When using only a single value the network did not improve over the performance found for the task with only a single input sample (i.e. the same as in Fig. 5.4). The other way around using exponentially distributed values did decrease the performance in the delayed response task. In general, results did not change when using larger number of neurons and smaller or larger input populations (for all these additional parameter settings, see appendix A.

5.4 Discussion

After showing the potential of reward-mediated learning to lead to the development of cue integration abilities close to those predicted by an optimal Bayesian observer, we showed possible ways how this could be implemented in a more detailed biologically plausible structure. In a similar attempt Pfeiffer and colleagues [Pfeiffer *et al.* 2010] previously proposed a feed-forward model that showed approximate Bayesian inference with the use of reward-mediated Hebbian learning. To get to those results the authors required a specially prepared input state space, that essentially was a form of probabilistic population code and thereby still required probabilistic computations as intrinsic feature of the model. Additionally they used a specially designed learning rule for which there is no biological evidence so far. We use a recurrent neural network with binary threshold neurons, which is simple yet similar to the spiking of real neurons and the connectivity pattern of the brain. Previous work on recurrent network showed their potential to approximate non-linear functions [Maass *et al.* 2003, Maass *et al.* 2006]. Usually one has to exhaustively search the parameter space before finding a network that is able to perform a certain approximation. Our focus in this work is on the plasticity mechanisms that could adapt the parameters based on the input and help the network discover the means of integrating information from two noisy cues.

In comparison to previous work on plasticity in very similar networks [Lazar *et al.* 2007, Lazar *et al.* 2009, Savin 2010], we test our model in a more complex task, involving noisy stimuli. These and other studies could already show the usefulness of a number of mechanisms, namely IP, synaptic scaling and STDP for improvements in the performance of the networks. We add another form of plasticity on the inhibitory synapses, inspired by biological data [Castillo *et al.* 2011] and some early theoretical work [Bourjaily & Miller 2011b, Sprekeler *et al.* 2011, Vogels *et al.* 2011]. For the output training, which is usually done in a supervised fashion [Lukoševičius & Jaeger 2009], we use a reward-modulated version of STDP, building on results from [Savin 2010], but using the TD error instead of an absolute reward signal to more closely match biological results.

In a task similar to the one used in Chapter 3, with an additional temporal delay between input and action, we show that a simple reservoir network can learn to perform better than either of the single cues. This shows that the model starts to integrate the two cues, although it is not as good as would be predicted from optimal use of all available information. One possible reason for the performance difference is the memory limitation imposed by using binary spiking neurons (see [Lazar *et al.* 2007, Savin 2010]). As we could see the performance of the network declines with increasing delay. On the other hand additional timesteps could increase the specificity of the reservoir state and make it easier for the linear output units to decode. We find some hints for this in the increasing selectivity of neurons for the causal model as time proceeds.

Because most previous results with reservoir networks were shown using inputs at multiple time points (e.g. in time-series prediction) [Lukoševičius & Jaeger 2009], we also test the network in such a task. In our accumulation task the network at every timestep receives a noisy sample of the input position. In this task it is possible to not only integrate the two cues, but also to temporally integrate the samples to improve performance. When comparing the performance with both single- and multi-step Bayesian observers, it seems that the network does not use the additional information. But if we compare the absolute performance to the delayed response task, which provides only single-step information, we find an improvement. The same is true for a comparison with the performance after learning a unisensory task. From these observations we can see that there is both temporal and multisensory integration happening in the reservoir. Both processes are far from optimal, information in the memory seems to be easily dominated by new input activation, and the integration of the two cues uses mostly information contained in the activity of the input population, which limits the performance. Nevertheless, we also think that this could be improved by using more powerful neuron models. At the same time we showed that the plasticity mechanisms used here can support extraction of information from both temporal and spatial sources.

For the training of the read-out weights, we could demonstrate that reward-mediated and biologically plausible online learning can result in performances similar to those reached by conventional supervised offline methods. As was also demonstrated in a few other recent publications [Legenstein *et al.* 2008, Savin 2010] an STDP rule that is modulated by a reward signal [Izhikevich 2007] is able to solve complex tasks. We could also show that by using the TD error instead of the absolute reward as modulating signal we can drop the synaptic scaling of the output weights without a loss in performance. For longer delays the performance of the R-STDP read-out falls behind the supervised method, which we believe is mostly due to the use of the homeostatic mechanism (eq (5.7)) in the output units of the former. We set the parameters so that every output unit will adapt its thresholds in a way that leads to uniform firing frequencies of all outputs over all trials. Although this matches the uniform distribution of input positions, it is not necessarily the pattern optimizing reward due to the non-linear border effects (see also Fig. A.4 in the appendix). We use that specific rule because it can control the timepoint of desired firing within a trial and at the same time prevent silent output units (e.g. due to the initialization). In [Savin 2010] a similar effect was reached by a combination of IP and R-STDP learning for the "go"-weight. We did not get good results when using this combination, specifically for temporal delays between the cues and in the accumulation task. We did though get similar results to the ones shown above when using IP only based on the firing in the action timestep. In general we think that all these plasticity rules are more biologically plausible than the screening for an optimal initialization or the use of completely different neuron types in the output.

Our results require a plastic reservoir, it is not sufficient to simply train a read-out unit on a random network. We show that the combination of intrinsic and inhibitory plasticity mechanisms can drive the reservoir to a regime where it can encode the information necessary to solve the task. As a result of IP, the number of neurons that are completely silent will be close to zero, which does improve the encoding capacities of the network [Schrauwen *et al.* 2008, Steil 2007]. The iSTDP rule that we used was previously shown to balance excitation and inhibition in a network of neurons [Sprekeler *et al.* 2011, Vogels *et al.* 2011]. Our results can, although we use much simpler neuron models, be interpreted in a similar way. Without iSTDP we find alternating timesteps that are dominated either by excitation or inhibition, leading to a very low sparseness in population activity, despite the presence of high lifetime sparseness of each neuron enforced by IP. When describing the effects of iSTDP on recurrent networks one should therefore also highlight its potential in generating such population sparsity, which is beneficial for encoding. In addition we also showed that, similar to an alternative inhibitory plasticity mechanism [Bourjaily & Miller 2011b, Bourjaily & Miller 2011a], iSTDP can increase the selectivity to pairs of inputs.

In contrast to previous studies [Lazar *et al.* 2007, Lazar *et al.* 2009, Savin 2010], we do not find further improvements when adding STDP or R-STDP. The difference to older work is a larger state space and more importantly noise in the input. To make up for this we had to increase the number of neurons in the reservoir and with the same connection probability also the number of synapses for each neuron. At least concerning R-STDP this could explain the absent impact, since it has been suggested that increasing the number of units decreases the average correlation of a unit's firing with the output action (a credit assignment problem), when using simple reward-dependent trace learning in spiking neuron networks [Urbanczik & Senn 2009]. These results apply probably in an even stronger way to a rule where STDP is modulated by the TD error, since this is a real error signal. In accordance with that we find worse performance for this rule compared to absolute reward modulation. Urbanczik and colleagues proposed an additional modulatory signal based on the correlation of each neurons firing with the average population activity, but this explicitly requires spike rate coding by the neurons, which we do not have in our model. Concerning STDP, our results are similar to another study testing for the effects of inhibitory plasticity in a network [Bourjaily & Miller 2011b], where it was found that the performance with STDP in combination with their inhibitory mechanism ("LTPi") was better than pure STDP but slightly worse than LTPi alone. The authors hypothesize that STDP is decreasing the diversity of connection patterns and therefore will only allow a limited set of selectivities. Despite their simpler task, we think the effects of the plasticities could be similar to what we find.

To conclude we showed that the combination of biologically plausible plasticity rules with simple spiking neurons in a recurrent network are able to learn a cue integration and causal inference task. We think that this is a promising result in the search for possible implementations of the theory proposed in Chapter 3, that reward-mediated learning could lead to near-optimal cue integration. The computational complexity of the units used here is much lower compared to those used in the more abstract neural network from this previous chapter. We also showed the positive effect of using multiple different plasticity mechanisms, in particular the combination of IP and iSTDP seems to have a large positive influence on the encoding capabilities of the reservoir. Future work will have to address both the cue integration performance and the positive plasticity effects in a network with more powerful neurons, for example in LSMs.

6 Conclusions and Future Work

In the last years many experimental studies could show that human performance in a variety of psychophysical tasks can match relatively closely the predictions of an optimal theoretical model [Alais & Burr 2004, Battaglia *et al.* 2003, Ernst & Banks 2002, Jacobs 1999, Knill & Saunders 2003, Mamassian & Landy 2001]. Attempts to model the underlying principles that are used by the brain seem to almost exclusively focus on Bayesian mechanisms encoded in the biological substrates [Deneve 2008a, Johnston *et al.* 1993, Ma *et al.* 2006, Rao 2004, Vul *et al.* 2009]. Despite some promising theories that came out of those studies, recent findings suggest that this can not be the full story. In cue integration, which is the most common task used in those experimental studies mentioned above, there seems to be a gradual increase in performance during development, which only in the end shows close similarities with the optimal predictions [Gori *et al.* 2008, Nardini *et al.* 2008, Nardini *et al.* 2010, Neil *et al.* 2006]. There are also a number of other questions regarding Bayesian mechanisms in the brain that are yet unanswered [Fiser *et al.* 2010, Rothkopf *et al.* 2010]. In this thesis I show that there are indeed alternatives to those explicit Bayesian methods, that nevertheless can lead to near-optimal performance. We take ideas from model-free reinforcement learning (RL), where behaviour is adapted only based on state-action-outcome pairings. Using this our model learns to perform a cue integration task without any prior knowledge about environmental structure or statistics. We describe the performance of this model on simulated data and compare it to the prediction of Bayesian models (Chapter 3). Interestingly, this model does not only learn to solve the cue integration task but also a more complex setup, where it additionally has to find out in which cases it is best to integrate and in which it is not (causal inference).

Those results were repeated on realistic data in Chapter 4. The RL model was able to perform two different multi-cue depth estimation tasks better than what was found with a standard approximation of Bayesian inference. The reason for this lies in various unknown correlations between and non-uniformities within the different cues used in this task. This shows the potential of the approach for robotics or computer science applications. It seems to be particularly interesting for the field of autonomous robotics, where structural knowledge about the environment is usually hard to obtain.

In Chapter 5 we test a possible neuronal implementation of the RL principle. We show that in a simplified model of a spiking neural network with recurrent connections the combination of a number of biologically plausible plasticity mechanisms can lead to the development of cue integration abilities. Despite a performance below the optimal predictions, possible reasons for which are discussed in section 5.4, we want to stress the potential of common (and also new) plasticity mechanisms that arises when they are put to work together. Consistent with previous work from our group [Lazar *et al.* 2007, Lazar *et al.* 2009, Lazar *et al.* 2011, Savin *et al.* 2010, Savin 2010, Triesch 2005b, Triesch 2007], our results show that using multiple plasticity rules at the same time can greatly improve the performance of even simple networks over untrained ones, or those that only use a subset of plasticities. Specifically, the RL principles of learning in interaction can be adapted to the implementational level and lead to successful learning of cue integration. We believe that slightly more detailed neuron models like those used in liquid-state machines (LSMs)/echo state networks (ESNs) can also benefit from this principles (see e.g. [Savin *et al.* 2010, Triesch 2007]) and may be able to narrow the current gap between network performance and Bayesian predictions.

Despite a promising overall results of our RL-based model, we do not propose this system as replacement of model-based or Bayesian theories. We are well aware that many experimental results clearly point towards the explicit use of probability computations and knowledge of task structure (e.g. [Atkins *et al.* 2001, Mamassian & Landy 2001, Vul *et al.* 2009]) and therefore support the existence of those types of models in the brain. Instead we want to raise awareness of alternative ideas that could also be able to explain some of the experimental results and especially want to emphasize the importance of developmental processes and learning that could lead to the emergence of whatever type of model. In addition the multiple controller hypothesis (as introduced in section 2.3.3) states a clear need for research on both model-free and model-based theories. Our work shows that also for more complex tasks the habitual controller will learn to perform as good as the goal-directed one and could therefore take over his duties with a benefit in, e.g., processing time.

An alternative view on a multiple controller system for cue integration can arise if we follow the arbitration mechanism proposed in [Keramati *et al.* 2011]. The orienting task that we use as an example in our simulations (but also, e.g., the depth estimation tasks from Chapter 4) have to be performed fast and frequently in everyday life, therefore their reward rate can be expected to be relatively high. If this is true, the model-free controller would be chosen already very early in the learning process, which could explain the gradual development of cue integration abilities and also the near-optimal performance in older individuals. For fast changing environmental statistics like those used for example in [Triesch *et al.* 2002] the system would choose the model-based system, which could explain the fast adaptation of human behaviour. But as also shown in Chapter 3, our model-free system is able to adapt within a small number of trials to those changes as well (given they are not too large), and could therefore soon take back control.

Lastly, despite some theoretical proposals [Beck *et al.* 2008, Deneve 2008a, Ma *et al.* 2006] of how Bayesian computations could be implemented and how learning in those systems could take place [Deneve 2008b, Pfeiffer *et al.* 2010], there is by now still a lack of experimental validation for those models. In contrast the neuron models and plasticity mechanisms that are used in Chapter 5 are all based on neurobiological data (although abstracted). For this reason I think more research has to be done about the potential of existing mechanisms in generating the observed behaviour. In addition work on Bayesian methods should also search for theories that could include the developmental aspects of human behaviour. This thesis does try to contribute to the further issue and can hopefully result in further advancement of the knowledge of cue integration, RL and the interaction and potential of different neuronal plasticity mechanisms.

Future work

The results in Chapter 4 using realistic data were computed offline, and we could easily compute a reward signal by using the knowledge about the true depth for each input. A next step would be to include the algorithm in a real robotic system. In a depth recognition task similar to what we used in Chapter 4 the robot could do a grasping movement and use the success as a reward signal. This reward signal would be noisy, because it e.g. also depends on the hand position, and it would be interesting to see how the model deals with that. An additional difficulty would come from using cluttered scenes, although this will first of all affect the estimates of the individual cues but not their integration. In general we think that this method will be widely usable, given that one finds good ways to compute reward signals for each application.

A second topic that could be researched building on the results from this thesis is the interplay between model-based and model-free learning. It is particularly interesting how the fact that one mechanism determines the policy impacts the learning of the second one. Work on the multiple controller theory was so far usually using parallel state sampling approaches [Daw *et al.* 2005, Keramati *et al.* 2011, Lengyel & Dayan 2008]. Most RL algorithms are only guaranteed to converge to the optimal solution for specific action-selection rules that allow for enough exploration. The model-based controller should therefore select the actions using e.g. a softmax policy (that is doing some exploration) to allow the model-free system to also learn from the outcomes. It would be interesting to see if, in that case, a decreasing temparature variable might partially explain the gradual improvement of children's cue integration.

The research on the effect of multiple plasticity mechanisms is still in its infancy. The biggest limitation of our work and that of others currently is the use of very abstract and simplified neuron models. It would be of great interest to test if these positive results of interaction could also be found in LSMs or similar more complex networks. Importantly, the memory potential of the reservoir in those networks is much larger. For that reason we could hope to really profit from the separation ability of the network and potentially find an increase in performance with delay time. In a continuous time reservoir there are also many more opportunities to test the influence of delays between two cues on causal inference processes. The plasticity rules we are using are also originally found/designed on those more complex neuron models and for example intrinsic plasticity (IP) might not only change the threshold but also the gain function [Triesch 2005a]. On the other hand, we did deliberately choose very simple dynamics to limit the number of free parameters which tend to explode when using multiple plasticity mechanisms in complex neuron models.

A research topic that only recently started to grow is inhibitory plasticity. Based on a variety of experimental results only in 2011 the first theoretical plasticity rules were proposed [Bourjaily & Miller 2011b, Sprekeler *et al.* 2011] with some adaptations or applications of existing rules (standard spike-time-dependent plasticity (STDP), IP, inverse STDP, etc) also mentioned in the literature [Castillo *et al.* 2011, Feldman 2009, Maffei 2011]. The effects of those mechanisms in neural networks, especially in interaction with excitatory plasticity, requires further investigation. Since the diversity among inhibitory neurons in the brain is much larger than that of excitatory ones [Markram *et al.* 2004], it will also be interesting to combine multiple inhibitory plasticity mechanisms in the same network. In most reservoir networks the inhibitory population only acts to maintain a certain balance in excitation, neglecting its potential to participate in encoding, like e.g. in surround inhibition [Hirsch 2003] or predictive coding [Rao & Ballard 1999]. We therefore think that inhibitory plasticity has the potential to have a big impact on the computational power of those networks.

Finally, it would be interesting to research the potential of existing models like probabilistic population codes in combination with reservoir networks and/or with low-level plasticity mechanisms like those we used in Chapter 5. One would have to use spiking neurons with a stochastic component [Gerstner & Kistler 2002], which were recently shown to be suitable for reservoir computing [Schliebs *et al.* 2011]. The hope would be that after learning the network would be able to compute with probabilities but that the relation between the neurons and the implicit meaning of population firing would only arise due to neuronal plasticity mechanisms.

Appendix

A
Appendix: Recurrent Neural Networks

In the following we will show additional results from our recurrent network experiments, that we got while varying different aspects of the model. These results do not impact the main points that we made in Chapter 5, but we think they might still be useful for future work on similar models. We show the influence of changing some of the variables on performance and dynamics of the network to also provide some hints about potential ways to improve those networks. This is by no means an exhaustive screening of the parameter space, but more of a selected, while sometimes hopefully exemplary, testing of robustness and sensitivity with respect to certain dimensions of this space. All the results on these pages apply to the delayed response task.

A.1 Initialization/Network Structure

In Fig. A.1 we show the average firing patterns after training the reservoir without IP for a different initialization of the excitatory threshold θ^{e} (Compare Fig. 5.12 in Chapter 5). In Fig. 5.12 we set $\theta^{\mathrm{e}} = 0.1$ for all neurons and we wanted to test how much our results were influenced by that choice. We therefore tested for a reservoir with Gaussian distributed values covering the range seen after training with IP ($\mathcal{N}(0.3, 0.08)$, Fig. A.1), resulting in stereotyped firing relatively independent of the input and many silent neurons. The performance in this setting is also shown in Fig. A.3**D**.

We did not find differences in performance compared to the standart setting from e.g. Fig. 5.4 for a number of variations in other parameters: $|Z| = 10, 30$, $p_{ee} = 0.03$ Explicitly choosing a subset of excitatory reservoir neurons to not receive any connections from input neurons only decreased performance. The same can be seen for larger connection probabilities ($p_{ee} = 0.5$, Fig. A.3**E**). Increasing the number of reservoir neurons to ($N_e = 800$, $N_i = 200$) or ($N_e = 1000$, $N_i = 250$) without changing any other parameters did also decrease the performance.

A.2 Plasticities

Figure A.2 shows how IP and iSTDP influence the sparsity of the reservoir activity. Without iSTDP all neurons adapt their firing thresholds to very closely match the target firing rate θ^{e}. But in return this leads to distributions of population activity that differ much between the timepoints after the input. On the other hand, turning off IP shifts the average firing rate of the neurons to much larger numbers, although the exact values depend on the initial setting of the thresholds (see also Fig. A.1). In addition the distribution of population activity gets flatter and also covers a larger space of values for the first step after the input (blue), and is largely 0 thereafter.

When we only use iSTDP in half of the inhibitory population, while the rest keeps its initial weights,

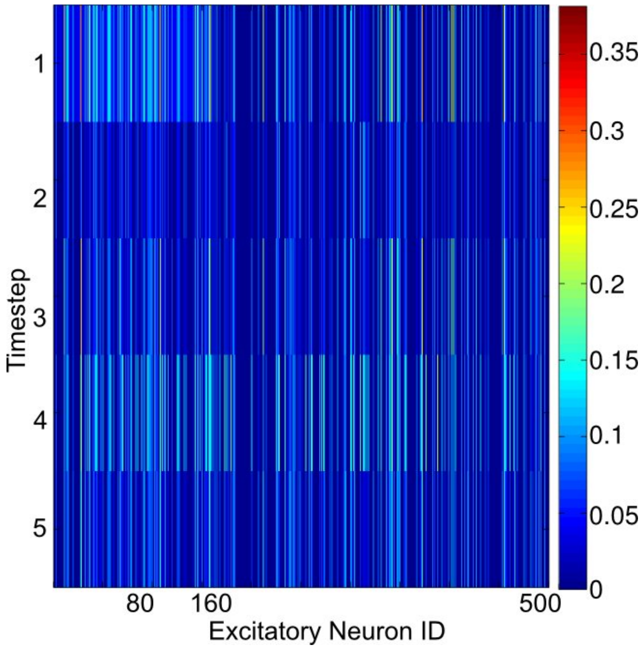

Figure A.1. Average firing of the reservoir neurons over all inputs without IP. The figure shows the firing of each excitatory neuron in the reservoir averaged over 10,000 stimuli. The network is using STDP, iSTDP and scaling, but no IP and is initialized with $\theta^{e_i} = \mathcal{N}(0.3, 0.08)$. The mean value was chosen to be close to the mean of the final thresholds in a network using IP. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$, $dT = 2$).

results are not different from a network without any inhibitory plasticity (Fig. A.3**F**)). When looking at the final weights of the plastic neurons after training, we find that they all decayed to zero.

We also run the training with different values for the target firing rate λ^{e} of the IP rule. Figure A.3 shows the networks performance for $\lambda^{\mathrm{e}} = 0.0082$ (**B**), $\lambda^{\mathrm{e}} = 0.029$ (**C**) (compare with standart setting of $\lambda^{\mathrm{e}} = 0.021$) and λ^{e} drawn from the exponential distribution $exp(\frac{\nu_{\mathrm{inp}}}{N_e}) + 0.005$) (**A**) except for input neurons which use $\lambda^{\mathrm{e}} = \frac{1}{|Z|}$ (that is the same setting as was used for the accumulation task setup).

A.3 Read-out

As mentioned in Section 5.4 we find that the optimal distribution of outputs over all possible input states is not uniform, as could have been expected from the uniform distribution of the input positions (Fig. A.4). This is due to the border effects, where the noise can drive stimuli to the outside of what is seen by the reservoir. It can be expected that the fact that we actually use a uniform target firing rate for the homeostatic plasticity of the output units influences the performance of the model.

We tested the potential of read-out neurons with a slight variation on the winner-take-all (WTA) mechanism. Instead of choosing the neuron with highest potential among all active units, we chose the one

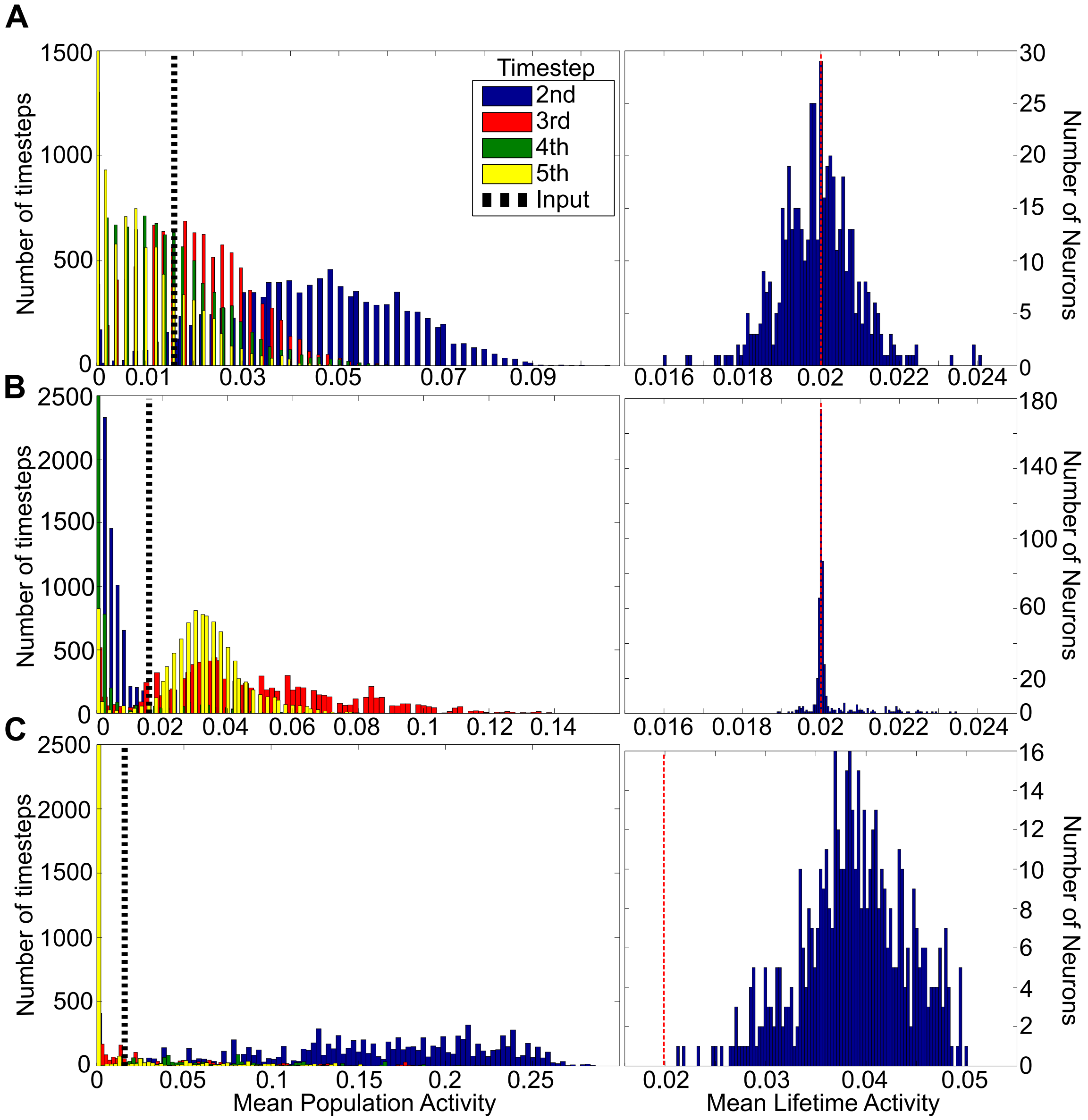

Figure A.2. Population and Lifetime Sparsity for different plasticity mechanisms. This figure shows the mean activity of the reservoir over 10,000 trials depending on which plasticity rules are active. The left column shows the mean population activity in each timestep after an input, displaying the population sparsity of the reservoir. The dashed black line denotes the mean population activity caused directly by the input (in step one). The right column is plotting a histogram of the mean activity of each single neuron over all timesteps to evaluate lifetime sparsity. The dashed red line highlights the target firing rate θ^{e} of the IP rule. **A**: All plasticities. **B**: STDP, scaling and IP but no iSTDP. **C**: STDP, scaling and iSTDP but no IP. (in all cases $\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$, $dT = 2$).

with highest relative surpassing of its threshold:

$$z'_j(t+1) = \begin{cases} 1, & \text{if } z'^{\text{pot}}_j(t+1) > \theta_j \text{ AND } z'^{\text{pot}}_j(t+1) = \max\limits_l\left(\frac{z'^{\text{pot}}_l(t+1)-\theta_l}{\theta_l}\right) \\ 0, & \text{else} \end{cases} \tag{A.1}$$

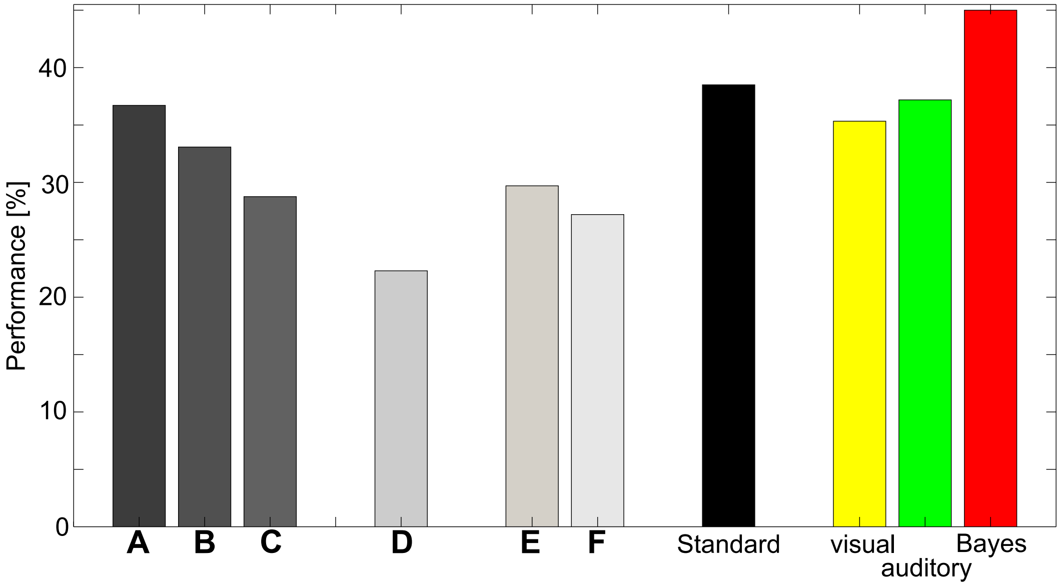

Figure A.3. Performance of the network with different initializations. The figure shows the performance of the RL read-out for networks with different initializations of certain variables (**A**-**F**). For comparison we also plot the performance of the unisensory observers and Bayesian integration as well as the result we got with our standart setting from Chapter 5. **A**, **B** and **C** show results from varying λ^{e} (exponentially distributed, low and high values respectively). **D** is the performance without IP and with Gaussian distributed thresholds θ^{e}. In **E** excitatory connectivity is very dense ($p_{ee} = 0.5$). A network where only half of the inhibtory neurons use iSTDP is producing the reward shown in **F**. ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$, $dT = 2$).

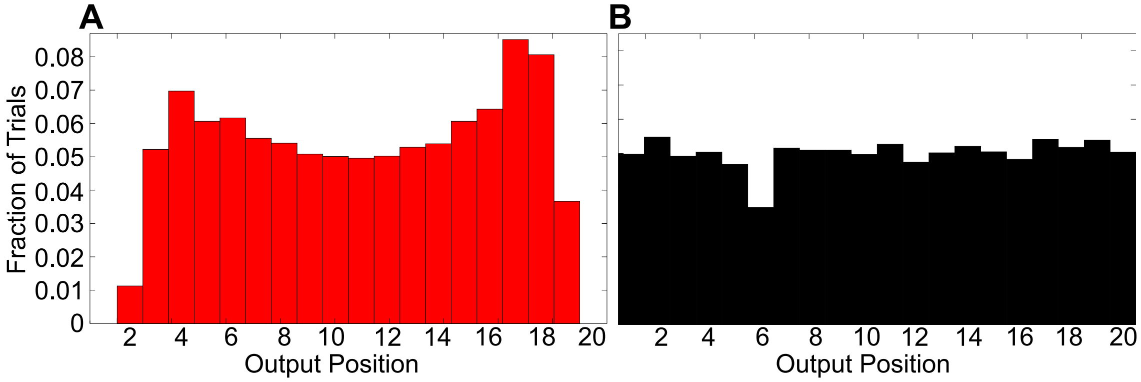

Figure A.4. Distribution of output firing. Histograms of the fraction of trials that each output unit fires averaged over all input positions. Comparison between the Bayesian observer (**A**) and the network with homeostatic plasticity (eq (5.7)) targeting a uniform distribution (**B**). ($\sigma_a^2 = 3$, $\sigma_v^2 = 3.2$, $p(C = 1) = 0.5$, $dT = 2$).

This could be seen as more plausible with respect to neuronal implementations of WTA that are based on firing rates, which in turn will depend on the membrane potential relative to a neuron's threshold. The results with this alternative decision rule did not change compared to the method used in Chapter

5.

Bibliography

[Adams & Dickinson 1981] Christopher D Adams and Anthony Dickinson. *Instrumental responding following reinforcer devaluation.* Quarterly Journal of Experimental Psychology B: Comparative and Physiological, vol. 33B, no. 2, pages 109–121, 1981.

[Adams 1982] Christopher D Adams. *Variations in the sensitivity of instrumental responding to reinforcer devaluation.* Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology, vol. 34, pages 77–98, 1982.

[Adler & Tso 1974] Julius Adler and Wung-Wei Tso. *"Decision"-making in bacteria: chemotactic response of Escherichia coli to conflicting stimuli.* Science, vol. 184, no. 143, pages 1292–4, June 1974.

[Alais & Burr 2004] David Alais and David Burr. *The Ventriloquist Effect Results from Near-Optimal Bimodal Integration.* Current Biology, vol. 14, no. 3, pages 257–262, February 2004.

[Anastasio & Patton 2003] Thomas J Anastasio and Paul E Patton. *A Two-Stage Unsupervised Learning Algorithm Reproduces Multisensory Enhancement in a Neural Network Model of the Corticotectal System.* Journal of Neuroscience, vol. 23, no. 17, pages 6713–6727, July 2003.

[Anastasio *et al.* 2000] Thomas J Anastasio, Paul E Patton and Kamel Belkacem-Boussaid. *Using Bayes' rule to model multisensory enhancement in the superior colliculus.* Neural Computation, vol. 12, no. 5, pages 1165–1187, May 2000.

[Atiya & Parlos 2000] Amir F Atiya and Alexander G Parlos. *New results on recurrent network training: unifying the algorithms and accelerating convergence.* IEEE Transactions on Neural Networks, vol. 11, no. 3, pages 697–709, January 2000.

[Atkins *et al.* 2001] Joseph E Atkins, József Fiser and Robert A Jacobs. *Experience-dependent visual cue integration based on consistencies between visual and haptic percepts.* Vision Research, vol. 41, no. 4, pages 449–461, February 2001.

[Atkins *et al.* 2003] Joseph E Atkins, Robert A Jacobs and David C Knill. *Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts.* Vision Research, vol. 43, no. 25, pages 2603–2613, November 2003.

[Avillac *et al.* 2007] Marie Avillac, Suliann Ben Hamed and Jean-René Duhamel. *Multisensory integration in the ventral intraparietal area of the macaque monkey.* Journal of Neuroscience, vol. 27, no. 8, pages 1922–1932, February 2007.

[Baddeley *et al.* 1997] Roland Baddeley, Larry F Abbott, Michael C Booth, Frank Sengpiel, Tobe Freeman, Edward A Wakeman and Edmund T Rolls. *Responses of neurons in primary and inferior temporal visual cortices to natural scenes.* Proceedings of the Royal Society B: Biological Sciences, vol. 264, no. 1389, pages 1775–83, December 1997.

[Bahrick *et al.* 2002] Lorraine E Bahrick, Ross Flom and Robert Lickliter. *Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants.* Developmental Psychobiology, vol. 41, no. 4, pages 352–363, December 2002.

[Bair *et al.* 2007] Woei-Nan Bair, Tim Kiemel, John J Jeka and Jane E Clark. *Development of multisensory reweighting for posture control in children.* Experimental Brain Research, vol. 183, no. 4, pages 435–446, December 2007.

[Balleine & Dickinson 1998] Bernard W Balleine and Anthony Dickinson. *Goal-directed instrumental action: contingency and incentive learning and their cortical substrates.* Neuropharmacology, vol. 37, no. 4-5, pages 407–419, April 1998.

[Bao *et al.* 2001] Shaowen Bao, Vincent T Chan and Michael M Merzenich. *Cortical remodelling induced by activity of ventral tegmental dopamine neurons.* Nature, vol. 412, no. 6842, pages 79–83, July 2001.

[Bartlett & Baxter 2000] Peter L Bartlett and Jonathan Baxter. *A Biologically Plausible and Locally Optimal Learning Algorithm for Spiking Neurons.* Rapport technique, Australian National University, 2000.

[Barutchu *et al.* 2009a] Ayla Barutchu, David P Crewther and Sheila G Crewther. *The race that precedes coactivation: development of multisensory facilitation in children.* Developmental Science, vol. 12, no. 3, pages 464–473, April 2009.

[Barutchu *et al.* 2009b] Ayla Barutchu, Jaclyn Danaher, Sheila G. Crewther, Hamish Innes-Brown, Mohit N. Shivdasani and Antonio G. Paolini. *Audiovisual integration in noise by children and adults.* Journal of Experimental Child Psychology, vol. 105, pages 38–50, October 2009.

[Battaglia *et al.* 2003] Peter W Battaglia, Robert A Jacobs and Richard N Aslin. *Bayesian integration of visual and auditory signals for spatial localization.* Journal of the Optical Society of America. A, Optics, image science, and vision, vol. 20, no. 7, pages 1391–1397, July 2003.

[Beck *et al.* 2008] Jeffrey M Beck, Wei Ji Ma, Roozbeh Kiani, Tim Hanks, Anne K Churchland, Jamie Roitman, Michael N Shadlen, Peter E Latham and Alexandre Pouget. *Probabilistic Population Codes for Bayesian Decision Making.* Neuron, vol. 60, no. 6, pages 1142–1152, December 2008.

[Beierholm *et al.* 2011] Ulrik R Beierholm, Cedric Anen, Steven Quartz and Peter Bossaerts. *Separate encoding of model based and model free valuations in the human brain.* NeuroImage, vol. 58, no. 3, pages 962–955, July 2011.

[Berglund & Sitte 2005] Erik Berglund and Joaquin Sitte. *Sound source localisation through active audition.* In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pages 653–658, Edmonton, Canada, 2005. IEEE.

[Berns *et al.* 2001] Gregory S Berns, Samuel M McClure, Giuseppe Pagnoni and P Read Montague. *Predictability Modulates Human Brain Response to Reward.* Journal of Neuroscience, vol. 21, no. 8, pages 2793–2798, 2001.

[Bi & Poo 1998] Guo-qiang Bi and Mu-ming Poo. *Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type.* Journal of Neuroscience, vol. 18, no. 24, pages 10464–72, December 1998.

[Bi & Poo 2001] Guo-qiang Bi and Mu-ming Poo. *Synaptic modification by correlated activity: Hebb's postulate revisited.* Annual Review of Neuroscience, vol. 24, pages 139–66, January 2001.

[Bienenstock *et al.* 1982] Elie L Bienenstock, Leon Cooper and Paul Munro. *Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex.* Journal of Neuroscience, vol. 2, no. 1, pages 32–48, January 1982.

[Binda *et al.* 2007] Paola Binda, Aurelio Bruno, David C Burr and Maria C Morrone. *Fusion of visual and auditory stimuli during saccades: a Bayesian explanation for perisaccadic distortions.* Journal of Neuroscience, vol. 27, no. 32, pages 8525–8532, August 2007.

[Binns & Salt 1996] KE Binns and TE Salt. *Importance of NMDA receptors for multimodal integration in the deep layers of the cat superior colliculus.* Journal of Neurophysiology, vol. 75, no. 2, pages 920–930, February 1996.

[Bishop 2006] Christopher M Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer, New York, 2006.

[Bliss & Lø mo 1973] Tim V P Bliss and Terje Lø mo. *Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path.* Journal of Physiology, vol. 232, no. 2, pages 331–56, July 1973.

[Bornstein & Daw 2011] Aaron M Bornstein and Nathaniel D Daw. *Multiplicity of control in the basal ganglia: computational roles of striatal subregions.* Current Opinion in Neurobiology, vol. 21, no. 3, pages 374–380, March 2011.

[Bourjaily & Miller 2011a] Mark A Bourjaily and Paul Miller. *Excitatory, Inhibitory, and Structural Plasticity Produce Correlated Connectivity in Random Networks Trained to Solve Paired-Stimulus Tasks.* Frontiers in Computational Neuroscience, vol. 5, page 37, 2011.

[Bourjaily & Miller 2011b] Mark A Bourjaily and Paul Miller. *Synaptic Plasticity and Connectivity Requirements to Produce Stimulus-Pair Specific Responses in Recurrent Networks of Spiking Neurons.* PLoS Computational Biology, vol. 7, no. 2, page e1001091, February 2011.

[Bradski & Kaehler 2008] Gary Bradski and Adrian Kaehler. Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Inc., October 2008.

[Brandwein *et al.* 2011] Alice B Brandwein, John J Foxe, Natalie N Russo, Ted S Altschuler, Hilary Gomes and Sophie Molholm. *The Development of Audiovisual Multisensory Integration Across Childhood and Early Adolescence: A High-Density Electrical Mapping Study.* Cerebral Cortex, vol. 21, no. 5, pages 1042–1055, May 2011.

[Braun *et al.* 2010] Daniel A Braun, Stephan Waldert, Ad Aertsen, Daniel M Wolpert and Carsten Mehring. *Structure learning in a sensorimotor association task.* PloS one, vol. 5, no. 1, page e8973, 2010.

[Brayanov & Smith 2010] Jordan B Brayanov and Maurice A Smith. *Bayesian and "Anti-Bayesian" Biases in Sensory Integration for Action and Perception in the Size-Weight Illusion.* Journal of Neurophysiology, vol. 103, no. 3, pages 1518–1531, March 2010.

[Bromberg-Martin *et al.* 2010] Ethan S Bromberg-Martin, Masayuki Matsumoto and Okihide Hikosaka. *Dopamine in Motivational Control: Rewarding, Aversive, and Alerting.* Neuron, vol. 68, no. 5, pages 815–834, 2010.

[Brown *et al.* 2003] Myron Z Brown, Darius Burschka and Gregory D Hager. *Advances in computational stereo.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pages 993–1008, August 2003.

[Bruce *et al.* 1981] Charles Bruce, Robert Desimone and Charles G Gross. *Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque.* Journal of Neurophysiology, vol. 46, no. 2, pages 369–84, August 1981.

[Bruns *et al.* 2011] Patrick Bruns, Charles Spence and Brigitte Röder. *Tactile recalibration of auditory spatial representations.* Experimental Brain Research, vol. 209, no. 3, pages 333–344, March 2011.

[Bryan & Harter 1897] William Lowe Bryan and Noble Harter. *Studies in the physiology and psychology of the telegraphic language.* Psychological Review, vol. 4, no. 1, pages 27–53, 1897.

[Bülthoff & Yuille 1991] Heinrich H Bülthoff and Alan L Yuille. *Bayesian Models for Seeing Shapes and Depth.* Comments on Theoretical Biology, vol. 2, no. 4, pages 283–314, 1991.

[Burge *et al.* 2010] Johannes Burge, Ahna R Girshick and Martin S Banks. *Visual-Haptic Adaptation Is Determined by Relative Reliability.* Journal of Neuroscience, vol. 30, no. 22, pages 7714–7721, June 2010.

[Burnett *et al.* 2007] Luke Burnett, Barry E Stein, Thomas J Perrault and Mark T Wallace. *Excitotoxic lesions of the superior colliculus preferentially impact multisensory neurons and multisensory integration.* Experimental Brain Research, vol. 179, no. 2, pages 325–338, May 2007.

[Butler *et al.* 2011] John S Butler, Jennifer L Campos, Heinrich H Bülthoff and Stuart T Smith. *The Role of Stereo Vision in VisualVestibular Integration.* Seeing and Perceiving, vol. 24, no. 5, pages 453–470, 2011.

[Calabresi *et al.* 1992] P Calabresi, R Maj, A Pisani, N B Mercuri and G Bernardi. *Long-term synaptic depression in the striatum: physiological and pharmacological characterization.* Journal of Neuroscience, vol. 12, no. 11, pages 4224–33, November 1992.

[Calabresi *et al.* 2007] Paolo Calabresi, Barbara Picconi, Alessandro Tozzi and Massimiliano Di Filippo. *Dopamine-mediated regulation of corticostriatal synaptic plasticity.* Trends in Neurosciences, vol. 30, no. 5, pages 211–9, May 2007.

[Caplin *et al.* 2010] Andrew Caplin, Mark Dean, Paul W Glimcher and Robb B Rutledge. *Measuring Beliefs and Rewards: A Neuroeconomic Approach.* Quarterly Journal of Economics, vol. 125, no. 3, pages 923–960, August 2010.

[Caporale & Dan 2008] Natalia Caporale and Yang Dan. *Spike timing-dependent plasticity: a Hebbian learning rule.* Annual Review of Neuroscience, vol. 31, pages 25–46, January 2008.

[Castillo *et al.* 2011] Pablo E Castillo, Chiayu Q Chiu and Reed C Carroll. *Long-term plasticity at inhibitory synapses.* Current Opinion in Neurobiology, vol. 21, no. 2, pages 328–338, February 2011.

[Chater *et al.* 2006] Nick Chater, Joshua B Tenenbaum and Alan L Yuille. *Probabilistic models of cognition: conceptual foundations.* Trends in Cognitive Sciences, vol. 10, no. 7, pages 287–91, July 2006.

[Cheng *et al.* 2007] Ken Cheng, Sara J Shettleworth, Janellen Huttenlocher and John J Rieser. *Bayesian integration of spatial information.* Psychological Bulletin, vol. 133, no. 4, pages 625–637, July 2007.

[Child & Wendt 1938] Irvin L Child and G R Wendt. *The temporal course of the influence of visual stimulation upon the auditory threshold.* Journal of Experimental Psychology, vol. 23, no. 2, pages 109–127, 1938.

[Colonius & Diederich 2004] Hans Colonius and Adele Diederich. *Multisensory interaction in saccadic reaction time: a time-window-of-integration model.* Journal of Cognitive Neuroscience, vol. 16, no. 6, pages 1000–1009, 2004.

[Colwill & Rescorla 1985] Ruth M Colwill and Robert A Rescorla. *Instrumental responding remains sensitive to reinforcer devaluation after extensive training.* Journal of Experimental Psychology: Animal Behavior Processes, vol. 11, no. 4, pages 520–536, 1985.

[Coppola *et al.* 1998] David M Coppola, H R Purves, A N McCoy and Dale Purves. *The distribution of oriented contours in the real world.* Proceedings of the National Academy of Sciences of the United States of America, vol. 95, no. 7, pages 4002–6, March 1998.

[Creese *et al.* 1983] Ian Creese, David R Sibley, Mark W Hamblin and Stuart E Leff. *The classification of dopamine receptors: relationship to radioligand binding.* Annual Review of Neuroscience, vol. 6, pages 43–71, January 1983.

[Crick 1989] Francis Crick. *The recent excitement about neural networks.* Nature, vol. 337, no. 6203, pages 129–32, January 1989.

[Cuppini *et al.* 2010] Cristiano Cuppini, Mauro Ursino, Elisa Magosso, Benjamin A Rowland and Barry E Stein. *An emergent model of multisensory integration in superior colliculus neurons.* Frontiers in Integrative Neuroscience, vol. 4, page 6, January 2010.

[Cuppini *et al.* 2011a] Cristiano Cuppini, Elisa Magosso and Mauro Ursino. *Organization, Maturation, and Plasticity of Multisensory Integration: Insights from Computational Modeling Studies.* Frontiers in Perception Science, vol. 2, page 77, May 2011.

[Cuppini *et al.* 2011b] Cristiano Cuppini, Barry E Stein, Benjamin A Rowland, Elisa Magosso and Mauro Ursino. *A computational study of multisensory maturation in the superior colliculus (SC).* Experimental Brain Research, vol. 213, no. 2, pages 341–349, September 2011.

[Dan & Poo 2004] Yang Dan and Mu-Ming Poo. *Spike timing-dependent plasticity of neural circuits.* Neuron, vol. 44, no. 1, pages 23–30, September 2004.

[Dankers *et al.* 2004] Andrew Dankers, Nick Barnes and Er Zelinsky. *Active vision - rectification and depth mapping.* In Australian Conference on Robotics and Automation. CiteSeerX - Scientific Literature Digital Library and Search Engine [http://citeseerx.ist.psu.edu/oai2] (United States), 2004.

[D'Ardenne *et al.* 2008] Kimberlee D'Ardenne, Samuel M McClure, Leigh E Nystrom and Jonathan D Cohen. *BOLD responses reflecting dopaminergic signals in the human ventral tegmental area.* Science, vol. 319, no. 5867, pages 1264–7, February 2008.

[Daw & Doya 2006] Nathaniel D Daw and Kenji Doya. *The computational neurobiology of learning and reward.* Current Opinion in Neurobiology, vol. 16, no. 2, pages 199–204, April 2006.

[Daw *et al.* 2005] Nathaniel D Daw, Yael Niv and Peter Dayan. *Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control.* Nature Neuroscience, vol. 8, no. 12, pages 1704–11, 2005.

[Daw *et al.* 2006] Nathaniel D Daw, John P O'Doherty, Peter Dayan, Ben Seymour and Raymond J Dolan. *Cortical substrates for exploratory decisions in humans.* Nature, vol. 441, no. 7095, pages 876–9, 2006.

[Daw *et al.* 2011] Nathaniel D Daw, Samuel J Gershman, Ben Seymour, Peter Dayan and Raymond J Dolan. *Model-Based Influences on Humans' Choices and Striatal Prediction Errors.* Neuron, vol. 69, no. 6, pages 1204–1215, 2011.

[Dayan & Daw 2008] Peter Dayan and Nathaniel D Daw. *Decision theory, reinforcement learning, and the brain.* Cognitive, Affective & Behavioral Neuroscience, vol. 8, no. 4, pages 429–453, December 2008.

[Dayan 1992] Peter Dayan. *The Convergence of TD(λ) for General λ.* Machine Learning, vol. 8, no. 3-4, pages 341–362, May 1992.

[de Winkel *et al.* 2010] Ksander N de Winkel, Jeroen Weesie, Peter J Werkhoven and Eric L Groen. *Integration of visual and inertial cues in perceived heading of self-motion.* Journal of Vision, vol. 10, no. 12, page 1, January 2010.

[Deneve 2008a] Sophie Deneve. *Bayesian spiking neurons I: inference.* Neural Computation, vol. 20, no. 1, pages 91–117, January 2008.

[Deneve 2008b] Sophie Deneve. *Bayesian spiking neurons II: learning.* Neural Computation, vol. 20, no. 1, pages 118–45, January 2008.

[Desai *et al.* 1999a] Niraj S Desai, Lana C Rutherford and Gina G Turrigiano. *BDNF regulates the intrinsic excitability of cortical neurons.* Learning & Memory, vol. 6, no. 3, pages 284–91, 1999.

[Desai *et al.* 1999b] Niraj S Desai, Lana C Rutherford and Gina G Turrigiano. *Plasticity in the intrinsic excitability of cortical pyramidal neurons.* Nature Neuroscience, vol. 2, no. 6, pages 515–20, June 1999.

[Dickinson 1985] Anthony Dickinson. *Actions and Habits: The Development of Behavioural Autonomy.* Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 308, no. 1135, pages 67–78, February 1985.

[Diederich & Colonius 2008] Adele Diederich and Hans Colonius. *Crossmodal interaction in saccadic reaction time: separating multisensory from warning effects in the time window of integration model.* Experimental Brain Research, vol. 186, no. 1, pages 1–22, March 2008.

[Ditterich 2010] Jochen Ditterich. *A Comparison between Mechanisms of Multi-Alternative Perceptual Decision Making: Ability to Explain Human Behavior, Predictions for Neurophysiology, and Relationship with Decision Theory.* Frontiers in Neuroscience, vol. 4, no. 184, 2010.

[Dolan 2007] Raymond J Dolan. *The human amygdala and orbital prefrontal cortex in behavioural regulation.* Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 362, no. 1481, pages 787–99, May 2007.

[Dorrn *et al.* 2010] Anja L Dorrn, Kexin Yuan, Alison J Barker, Christoph E Schreiner and Robert C Froemke. *Developmental sensory experience balances cortical excitation and inhibition.* Nature, vol. 465, no. 7300, pages 932–936, June 2010.

[Doya *et al.* 2007] Kenji Doya, Shin Ishii, Alexandre Pouget and Rajesh P N Rao. Bayesian Brain: Probabilistic Approaches to Neural Coding. MIT Press, Cambridge, UK, 2007.

[Drugowitsch *et al.* 2010] Jan Drugowitsch, Alexandre Pouget, Gregory C DeAngelis and Dora E Angelaki. *Optimal decision-making in multisensory integration.* In Computational and Systems Neuroscience (Cosyne 2010). Frontiers in Systems Neuroscience, 2010.

[D'Souza *et al.* 2010] Prashanth D'Souza, Shih-Chii Liu and Richard H R Hahnloser. *Perceptron learning rule derived from spike-frequency adaptation and spike-time-dependent plasticity.* Proceedings of the National Academy of Sciences of the United States of America, vol. 107, no. 10, pages 4722–4727, 2010.

[Elliott *et al.* 2009] Terry Elliott, Xutao Kuang, Nigel R Shadbolt and Klaus-Peter Zauner. *Adaptation in multisensory neurons: impact on cross-modal enhancement.* Network, vol. 20, no. 1, pages 1–31, 2009.

[Ernst & Banks 2002] Marc O Ernst and Martin S Banks. *Humans integrate visual and haptic information in a statistically optimal fashion.* Nature, vol. 415, no. 6870, pages 429–433, January 2002.

[Ernst *et al.* 2000] Marc O Ernst, Martin S Banks and Heinrich H Bülthoff. *Touch can change visual slant perception.* Nature Neuroscience, vol. 3, no. 1, pages 69–73, January 2000.

[Ernst 2004] Marc O Ernst. *Merging the senses into a robust percept.* Trends in Cognitive Sciences, vol. 8, no. 4, pages 162–169, April 2004.

[Ernst 2007] Marc O Ernst. *Learning to integrate arbitrary signals from vision and touch.* Journal of Vision, vol. 7, no. 5, pages 7.1–14, 2007.

[Farries & Fairhall 2007] Michael A Farries and Adrienne L Fairhall. *Reinforcement learning with modulated spike timing dependent synaptic plasticity.* Journal of Neurophysiology, vol. 98, no. 6, pages 3648–65, December 2007.

[Feldman 2009] Daniel E Feldman. *Synaptic mechanisms for plasticity in neocortex.* Annual Review of Neuroscience, vol. 32, pages 33–55, January 2009.

[Fetsch *et al.* 2009] Christopher R Fetsch, Amanda H Turner, Gregory C DeAngelis and Dora E Angelaki. *Dynamic Reweighting of Visual and Vestibular Cues during Self-Motion Perception.* Journal of Neuroscience, vol. 29, no. 49, pages 15601–15612, December 2009.

[Fiorillo *et al.* 2003] Christopher D Fiorillo, Philippe N Tobler and Wolfram Schultz. *Discrete coding of reward probability and uncertainty by dopamine neurons.* Science, vol. 299, no. 5614, pages 1898–902, March 2003.

[Fischer & Pena 2011] Brian J Fischer and Jose Luis Pena. *Owl's behavior and neural representation predicted by Bayesian inference.* Nature Neuroscience, vol. 14, no. 8, pages 1061–1066, July 2011.

[Fiser *et al.* 2010] József Fiser, Pietro Berkes, Gergö Orbán and Máté Lengyel. *Statistically optimal perception and learning: from behavior to neural representations.* Trends in Cognitive Sciences, vol. 14, no. 3, pages 119–130, March 2010.

[Florian 2007] RÄzvan V Florian. *Reinforcement Learning Through Modulation of Spike-Timing-Dependent Synaptic Plasticity.* Neural Computation, vol. 19, no. 6, pages 1468–1502, June 2007.

[Frémaux *et al.* 2010] Nicolas Frémaux, Henning Sprekeler and Wulfram Gerstner. *Functional Requirements for Reward-Modulated Spike-Timing-Dependent Plasticity.* Journal of Neuroscience, vol. 30, no. 40, pages 13326–13337, October 2010.

[Frens *et al.* 1995] Maarten A Frens, A John Van Opstal and Robert F Van der Willigen. *Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements.* Perception & Psychophysics, vol. 57, no. 6, pages 802–816, August 1995.

[Froemke & Dan 2002] Robert C Froemke and Yang Dan. *Spike-timing-dependent synaptic modification induced by natural spike trains.* Nature, vol. 416, no. 6879, pages 433–8, March 2002.

[Fujisaki *et al.* 2004] Waka Fujisaki, Shinsuke Shimojo, Makio Kashino and Shin'ya Nishida. *Recalibration of audiovisual simultaneity.* Nature Neuroscience, vol. 7, no. 7, pages 773–778, July 2004.

[Funahashi & Nakamura 1993] Ken-ichi Funahashi and Yuichi Nakamura. *Approximation of dynamical systems by continuous time recurrent neural networks.* Neural Networks, vol. 6, no. 6, pages 801–806, 1993.

[Fuzessery *et al.* 1985] Zoltan M Fuzessery, Jeffrey J Wenstrup and George D Pollak. *A representation of horizontal sound location in the inferior colliculus of the mustache bat (Pteronotus p. parnellii).* Hearing research, vol. 20, no. 1, pages 85–9, January 1985.

[Gabbott & Somogyi 1986] Paul L Gabbott and P Somogyi. *Quantitative distribution of GABA-immunoreactive neurons in the visual cortex (area 17) of the cat.* Experimental Brain Research, vol. 61, no. 2, pages 323–31, January 1986.

[Geisler 2011] Wilson S Geisler. *Contributions of ideal observer theory to vision research.* Vision Research, vol. 51, no. 7, pages 781–771, November 2011.

[Gepshtein *et al.* 2005] Sergei Gepshtein, Johannes Burge, Marc O Ernst and Martin S Banks. *The combination of vision and touch depends on spatial proximity.* Journal of Vision, vol. 5, no. 11, pages 1013–1023, 2005.

[Gershman *et al.* 2010] Sam Gershman, Edward Vul and Joshua B Tenenbaum. *Perceptual Multistability as Markov Chain Monte Carlo Inference.* In Y Bengio, D Schuurmans, J Lafferty, C K I Williams and A Culotta, editeurs, Advances in Neural Information Processing Systems 22 (NIPS 2009), Vancouver, Canada, 2010.

[Gerstner & Kistler 2002] Wulfram Gerstner and Werner M Kistler. Spiking Neuron Models: Single Neurons, Populations, Plasticity. Cambridge University Press, 1st édition, 2002.

[Gerstner *et al.* 1996] Wulfram Gerstner, Richard Kempter, J Leo van Hemmen and Hermann Wagner. *A neuronal learning rule for sub-millisecond temporal coding.* Nature, vol. 383, no. 6595, pages 76–81, September 1996.

[Gielen *et al.* 1983] S C Gielen, R A Schmidt and P J Van Den Heuvel. *On the nature of intersensory facilitation of reaction time.* Perception & Psychophysics, vol. 34, no. 2, pages 161–8, August 1983.

[Glimcher 2011] Paul W Glimcher. *Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. (Quantification of Behavior Sackler Colloquium).* Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. Supplement_3, pages 15647–15654, March 2011.

[Gold & Shadlen 2001] Joshua I Gold and Michael N Shadlen. *Neural computations that underlie decisions about sensory stimuli.* Trends in Cognitive Sciences, vol. 5, no. 1, pages 10–16, January 2001.

[Gold & Shadlen 2007] Joshua I Gold and Michael N Shadlen. *The neural basis of decision making.* Annual Review of Neuroscience, vol. 30, no. 1, pages 535–574, July 2007.

[Gopnik *et al.* 2004] Alison Gopnik, Clark Glymour, David M Sobel, Laura E Schulz, Tamar Kushnir and David Danks. *A theory of causal learning in children: causal maps and Bayes nets.* Psychological Review, vol. 111, no. 1, pages 3–32, 2004.

[Gori *et al.* 2008] Monica Gori, Michela Del Viva, Giulio Sandini and David C. Burr. *Young Children Do Not Integrate Visual and Haptic Form Information.* Current Biology, vol. 18, no. 9, pages 694–698, May 2008.

[Gori *et al.* 2010] Monica Gori, Giulio Sandini, Cristina Martinoli and David Burr. *Poor Haptic Orientation Discrimination in Nonsighted Children May Reflect Disruption of Cross-Sensory Calibration.* Current Biology, vol. 20, no. 3, pages 223–225, January 2010.

[Gottfried *et al.* 2003] Jay A Gottfried, John O'Doherty and Raymond J Dolan. *Encoding predictive reward value in human amygdala and orbitofrontal cortex.* Science, vol. 301, no. 5636, pages 1104–7, August 2003.

[Grant *et al.* 1951] David A Grant, Harold W Hake and John P Hornseth. *Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement.* Journal of Experimental Psychology, vol. 42, no. 1, pages 1–5, July 1951.

[Grunze *et al.* 1996] Heinz C Grunze, Donald G Rainnie, Michael E Hasselmo, Eddie Barkai, Elizabeth F Hearn, Robert W McCarley and Robert W Greene. *NMDA-dependent modulation of CA1 local circuit inhibition.* Journal of Neuroscience, vol. 16, no. 6, pages 2034–43, March 1996.

[Haeusler & Maass 2007] Stefan Haeusler and Wolfgang Maass. *A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models.* Cerebral Cortex, vol. 17, no. 1, pages 149–62, January 2007.

[Hairston *et al.* 2003] W David Hairston, Mark T Wallace, J William Vaughan, Barry E Stein, J L Norris and Jim A Schirillo. *Visual localization ability influences cross-modal bias.* Journal of Cognitive Neuroscience, vol. 15, no. 1, pages 20–29, January 2003.

[Harrar & Harris 2008] Vanessa Harrar and Laurence R Harris. *The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity.* Experimental Brain Research, vol. 186, no. 4, pages 517–524, April 2008.

[Hartman *et al.* 2006] Kenichi N Hartman, Sumon K Pal, Juan Burrone and Venkatesh N Murthy. *Activity-dependent regulation of inhibitory synaptic transmission in hippocampal neurons.* Nature Neuroscience, vol. 9, no. 5, pages 642–9, May 2006.

[Haruno & Kawato 2006] Masahiko Haruno and Mitsuo Kawato. *Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning.* Journal of Neurophysiology, vol. 95, no. 2, pages 948–59, February 2006.

[Hawkins *et al.* 1983] Robert D Hawkins, Thomas W Abrams, Thomas J Carew and Eric R Kandel. *A cellular mechanism of classical conditioning in Aplysia: activity-dependent amplification of presynaptic facilitation.* Science, vol. 219, no. 4583, pages 400–5, January 1983.

[Hayman & Eklundh 2002] Eric Hayman and Jan-Olof Eklundh. *Probabilistic and Voting Approaches to Cue Integration for Figure-Ground Segmentation.* In 7th European Conference on Computer Vision (ECCV 2002), volume 2352 of *Lecture Notes in Computer Science*, pages 469–486, Copenhagen, Denmark, April 2002. Springer.

[Hebb 1949] Donald O Hebb. The organization of behavior. Wiley-Interscience, New York, NY, USA, 1949.

[Held *et al.* 2011] Richard Held, Yuri Ostrovsky, Beatrice DeGelder, Tapan Gandhi, Suma Ganesh, Umang Mathur and Pawan Sinha. *The newly sighted fail to match seen with felt.* Nature Neuroscience, vol. 14, no. 5, pages 551–553, April 2011.

[Hershenson 1962] M Hershenson. *Reaction time as a measure of intersensory facilitation.* Journal of Experimental Psychology, vol. 63, pages 289–293, March 1962.

[Hikosaka *et al.* 2006] Okihide Hikosaka, Kae Nakamura and Hiroyuki Nakahara. *Basal ganglia orient eyes to reward.* Journal of Neurophysiology, vol. 95, no. 2, pages 567–84, 2006.

[Hillis *et al.* 2004] James M Hillis, Simon J Watt, Michael S Landy and Martin S Banks. *Slant from texture and disparity cues: optimal cue combination.* Journal of Vision, vol. 4, no. 12, pages 967–992, December 2004.

[Hirsch 2003] Judith A Hirsch. *Synaptic physiology and receptive field structure in the early visual pathway of the cat.* Cerebral Cortex, vol. 13, no. 1, pages 63–69, January 2003.

[Hoyer & Hyvärinen 2003] Patrik Hoyer and Aapo Hyvärinen. *Interpreting Neural Response Variability as Monte Carlo Sampling of the Posterior.* In Advances in Neural Information Processing Systems 15 (NIPS 2002), Vancouver, Canada, 2003.

[Hughes *et al.* 1994] Howard C Hughes, Patricia A Reuter-Lorenz, George Nozawa and Robert Fendrich. *Visual-auditory interactions in sensorimotor processing: saccades versus manual responses.* Journal of Experimental Psychology: Human Perception and Performance, vol. 20, no. 1, pages 131–153, February 1994.

[Huo & Murray 2009] Juan Huo and Alan Murray. *The adaptation of visual and auditory integration in the barn owl superior colliculus with Spike Timing Dependent Plasticity.* Neural Networks, vol. 22, no. 7, pages 913–921, September 2009.

[Hyvärinen *et al.* 2001] Aapo Hyvärinen, Juha Karhunen and Erkki Oja. Independent Component Analysis (Adaptive and Learning Systems for Signal Processing, Communications and Control). John Wiley & Sons, 1st édition, 2001.

[Izhikevich 2007] Eugene M Izhikevich. *Solving the distal reward problem through linkage of STDP and dopamine signaling.* Cerebral cortex, vol. 17, no. 10, pages 2443–52, October 2007.

[Jaakkola *et al.* 1994] Tommi Jaakkola, Michael I Jordan and Satinder P Singh. *On the Convergence of Stochastic Iterative Dynamic Programming Algorithms.* Neural Computation, vol. 6, no. 6, pages 1185–1201, November 1994.

[Jacobs & Fine 1999] Robert A Jacobs and I Fine. *Experience-dependent integration of texture and motion cues to depth.* Vision Research, vol. 39, no. 24, pages 4062–4075, December 1999.

[Jacobs & Kruschke 2011] Robert A Jacobs and John K Kruschke. *Bayesian learning theory applied to human cognition.* Wiley Interdisciplinary Reviews: Cognitive Science, vol. 2, no. 1, pages 8–21, 2011.

[Jacobs 1995] Robert A Jacobs. *Methods for combining experts' probability assessments.* Neural Computation, vol. 7, no. 5, pages 867–888, September 1995.

[Jacobs 1999] Robert A Jacobs. *Optimal integration of texture and motion cues to depth.* Vision Research, vol. 39, no. 21, pages 3621–3629, October 1999.

[Jaeger & Haas 2004] Herbert Jaeger and Harald Haas. *Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication.* Science, vol. 304, no. 5667, pages 78–80, April 2004.

[Jaeger 2001] Herbert Jaeger. *The echo state approach to analysing and training recurrent neural networks.* Rapport technique, German National Research Center for Information Technology, St. Augustin, Germany, 2001.

[Jaeger 2005] Herbert Jaeger. *Reservoir riddles: suggestions for echo state network research.* In International Joint Conference on Neural Networks (IJCNN 2005), volume 3, pages 1460–1462. IEEE, 2005.

[Jaime & Lickliter 2006] Mark Jaime and Robert Lickliter. *Prenatal exposure to temporal and spatial stimulus properties affects postnatal responsiveness to spatial contiguity in bobwhite quail chicks.* Developmental Psychobiology, vol. 48, no. 3, pages 233–242, April 2006.

[James 1890] William James. The principles of psychology. Henry Holt and Co, New York, NY, USA, 1890.

[Jay 2003] Thérèse M Jay. *Dopamine: a potential substrate for synaptic plasticity and memory mechanisms.* Progress in Neurobiology, vol. 69, no. 6, pages 375–90, April 2003.

[Jiang *et al.* 2007] Wan Jiang, Huai Jiang, Benjamin A Rowland and Barry E Stein. *Multisensory orientation behavior is disrupted by neonatal cortical ablation.* Journal of Neurophysiology, vol. 97, no. 1, pages 557–562, January 2007.

[Johnston *et al.* 1993] EB Johnston, BG Cumming and AJ Parker. *Integration of depth modules: stereopsis and texture.* Vision Research, vol. 33, no. 5-6, pages 813–826, 1993.

[Johnston *et al.* 1994] EB Johnston, BG Cumming and MS Landy. *Integration of stereopsis and motion shape cues.* Vision Research, vol. 34, no. 17, pages 2259–2275, September 1994.

[Junga *et al.* 1963] R Junga, Hans Helmut Kornhubera and J S Da Fonseca. *Multisensory Convergence on Cortical Neurons Neuronal Effects of Visual, Acoustic and Vestibular Stimuli in the Superior Convolutions of the Cat's Cortex.* Progress in Brain Research, vol. 1, pages 207–240, 1963.

[Kadunce *et al.* 1997] Daniel C Kadunce, J William Vaughan, Mark T Wallace, György Benedek and Barry E Stein. *Mechanisms of within- and cross-modality suppression in the superior colliculus.* Journal of Neurophysiology, vol. 78, no. 6, pages 2834–2847, December 1997.

[Kandel *et al.* 2000] Eric R Kandel, James H Schwartz and Thomas M Jessell. Principles of Neural Science. McGraw-Hill Medical, 4th édition, 2000.

[Karaoguz *et al.* 2010] Cem Karaoguz, Andrew Dankers, Tobias Rodemann and Mark Dunn. *An Analysis of Depth Estimation within Interaction Range.* In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), pages 3207–3212, Taipeh, Taiwan, October 2010. IEEE.

[Karaoguz *et al.* 2011] *Cem Karaoguz, *Thomas H Weisswange, *Tobias Rodemann, Britta Wrede and Constantin Rothkopf. *Reward-Based Learning of Optimal Cue Integration in Audio and Visual Depth Estimation.* In 15th International Conference on Advanced Robotics (ICAR 2011), Tallinn, Estonia, 2011. IEEE, *equal contribution authors.

[Kemp & Tenenbaum 2008] Charles Kemp and Joshua B Tenenbaum. *The discovery of structural form.* Proceedings of the National Academy of Sciences of the United States of America, vol. 105, no. 31, pages 10687–92, August 2008.

[Keramati *et al.* 2011] Mehdi Keramati, Amir Dezfouli and Payam Piray. *Speed/Accuracy Trade-Off between the Habitual and the Goal-Directed Processes.* PLoS Computational Biology, vol. 7, no. 5, page e1002055, May 2011.

[Kersten & Yuille 2003] Daniel Kersten and Alan L Yuille. *Bayesian models of object perception.* Current Opinion in Neurobiology, vol. 13, no. 2, pages 150–158, April 2003.

[Kersten *et al.* 2004] Daniel Kersten, Pascal Mamassian and Alan L Yuille. *Object Perception as Bayesian Inference.* Annual Review of Psychology, vol. 55, no. 1, pages 271–304, 2004.

[Khan & Shah 2001] Sohaib Khan and Mubarak Shah. *Object based segmentation of video using color, motion and spatial information.* In Conference on Computer Vision and Pattern Recognition (CVPR 2001), pages II–746 – II–751, Kauai, USA, 2001. IEEE Computer Society.

[Khan *et al.* 1995] Shahdid Khan, John L Spudich, James A McCray and David R Trentham. *Chemotactic signal integration in bacteria.* Proceedings of the National Academy of Sciences of the United States of America, vol. 92, no. 21, pages 9757–61, October 1995.

[Killcross & Coutureau 2003] Simon Killcross and Etienne Coutureau. *Coordination of Actions and Habits in the Medial Prefrontal Cortex of Rats.* Cerebral Cortex, vol. 13, no. 4, pages 400–408, April 2003.

[Kilman *et al.* 2002] Valerie Kilman, Mark C W van Rossum and Gina G Turrigiano. *Activity deprivation reduces miniature IPSC amplitude by decreasing the number of postsynaptic GABA(A) receptors clustered at neocortical synapses.* Journal of Neuroscience, vol. 22, no. 4, pages 1328–37, February 2002.

[Kirstein *et al.* 2008] Stephan Kirstein, Heiko Wersing and Edgar Körner. *A biologically motivated visual memory architecture for online learning of objects.* Neural Networks, vol. 21, no. 1, pages 65–77, January 2008.

[Knill & Pouget 2004] David C Knill and Alexandre Pouget. *The Bayesian brain: the role of uncertainty in neural coding and computation.* Trends in Neurosciences, vol. 27, no. 12, pages 712–719, December 2004.

[Knill & Richards 1996] David C Knill and Whitman Richards. Perception as Bayesian inference. Cambridge University Press, New York, New York, USA, 1996.

[Knill & Saunders 2003] David C Knill and Jeffrey A Saunders. *Do humans optimally integrate stereo and texture information for judgments of surface slant?* Vision Research, vol. 43, no. 24, pages 2539–2558, November 2003.

[Knill 2005] David C Knill. *Reaching for visual cues to depth: the brain combines depth cues differently for motor control and perception.* Journal of Vision, vol. 5, no. 2, pages 103–115, 2005.

[Knudsen & Brainard 1991] Eric I Knudsen and Michael S Brainard. *Visual instruction of the neural map of auditory space in the developing optic tectum.* Science, vol. 253, no. 5015, pages 85–87, July 1991.

[Körding *et al.* 2004] Konrad P Körding, Izumi Fukunaga, Ian S Howard, James N Ingram and Daniel M Wolpert. *A neuroeconomics approach to inferring utility functions in sensorimotor control.* PLoS Biology, vol. 2, no. 10, page e330, October 2004.

[Körding *et al.* 2007] Konrad P Körding, Ulrik R Beierholm, Wei Ji Ma, Steven Quartz, Joshua B Tenenbaum and Ladan Shams. *Causal inference in multisensory perception.* PLoS ONE, vol. 2, no. 9, page e943, 2007.

[Kullmann & Lamsa 2007] Dimitri M Kullmann and Karri P Lamsa. *Long-term synaptic plasticity in hippocampal interneurons.* Nature Reviews Neuroscience, vol. 8, no. 9, pages 687–99, September 2007.

[Kwakye *et al.* 2011] Leslie D Kwakye, Jennifer H Foss-Feig, Carissa J Cascio, Wendy L Stone and Mark T Wallace. *Altered auditory and multisensory temporal processing in autism spectrum disorders.* Frontiers in Integrative Neuroscience, vol. 4, no. 129, 2011.

[Lackner 1973] J R Lackner. *Visual rearrangement affects auditory localization.* Neuropsychologia, vol. 11, no. 1, pages 29–32, January 1973.

[Landy *et al.* 1995] Michael S Landy, Laurence T Maloney, Elizabeth B Johnston and Mark Young. *Measurement and modeling of depth cue combination: in defense of weak fusion.* Vision Research, vol. 35, no. 3, pages 389–412, February 1995.

[Laughlin *et al.* 1998] Simon B Laughlin, Rob R de Ruyter van Steveninck and John C Anderson. *The metabolic cost of neural information.* Nature Neuroscience, vol. 1, no. 1, pages 36–41, May 1998.

[Lazar *et al.* 2007] Andreea Lazar, Gordon Pipa and Jochen Triesch. *Fading memory and time series prediction in recurrent networks with different forms of plasticity.* Neural Networks, vol. 20, no. 3, pages 312–22, April 2007.

[Lazar *et al.* 2009] Andreea Lazar, Gordon Pipa and Jochen Triesch. *SORN: a self-organizing recurrent neural network.* Frontiers in Computational Neuroscience, vol. 3, no. October, page 23, January 2009.

[Lazar *et al.* 2011] Andreea Lazar, Gordon Pipa and Jochen Triesch. *Emerging Bayesian Priors in a Self-Organizing Recurrent Network.* In Timo Honkela, Wlodzislaw Duch, Mark Girolami and Samuel Kaski, editeurs, 21st International Conference on Artificial Neural Networks (ICANN 2011), volume 2 of *Lecture Notes in Computer Science*, pages 127–134, Espoo, Finland, 2011. Springer Berlin/Heidelberg.

[Lee *et al.* 1988] Choongkil Lee, William H Rohrer and David L Sparks. *Population coding of saccadic eye movements by neurons in the superior colliculus.* Nature, vol. 332, no. 6162, pages 357–60, March 1988.

[Legenstein *et al.* 2008] Robert Legenstein, Dejan Pecevski and Wolfgang Maass. *A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback.* PLoS Computational Biology, vol. 4, no. 10, page e1000180, October 2008.

[Lemon & Manahan-Vaughan 2006] Neal Lemon and Denise Manahan-Vaughan. *Dopamine D1/D5 receptors gate the acquisition of novel information through hippocampal long-term potentiation and long-term depression.* Journal of Neuroscience, vol. 26, no. 29, pages 7723–9, July 2006.

[Lengyel & Dayan 2008] Máté Lengyel and Peter Dayan. *Hippocampal contributions to control: the third way.* In J Platt, D Koller, Y Singer and S Roweis, editeurs, Advances in Neural Information Processing Systems 20 (NIPS 2007), pages 889–896. MIT Press, 2008.

[Leo *et al.* 2008] Fabrizio Leo, Caterina Bertini, Giuseppe di Pellegrino and Elisabetta Làdavas. *Multisensory integration for orienting responses in humans requires the activation of the superior colliculus.* Experimental Brain Research, vol. 186, no. 1, pages 67–77, March 2008.

[Lewald & Guski 2003] Jörg Lewald and Rainer Guski. *Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli.* Cognitive Brain Research, vol. 16, no. 3, pages 468–478, May 2003.

[Lewald 2002] Jörg Lewald. *Rapid adaptation to auditory-visual spatial disparity.* Learning & Memory, vol. 9, no. 5, pages 268–278, 2002.

[Lewkowicz & Ghazanfar 2006] David J Lewkowicz and Asif A Ghazanfar. *The decline of cross-species intersensory perception in human infants.* Proceedings of the National Academy of Sciences of the United States of America, vol. 103, no. 17, pages 6771–4, April 2006.

[Lewkowicz & Ghazanfar 2009] David J. Lewkowicz and Asif A. Ghazanfar. *The emergence of multisensory systems through perceptual narrowing.* Trends in Cognitive Sciences, vol. 13, no. 11, pages 470–478, November 2009.

[Lewkowicz & Turkewitz 1981] David J Lewkowicz and Gerald Turkewitz. *Intersensory interaction in newborns: modification of visual preferences following exposure to sound.* Child Development, vol. 52, no. 3, pages 827–32, September 1981.

[Lewkowicz 1996] David J Lewkowicz. *Perception of auditory-visual temporal synchrony in human infants.* Journal of Experimental Psychology: Human Perception and Performance, vol. 22, no. 5, pages 1094–1106, October 1996.

[Lukoševičius & Jaeger 2009] Mantas Lukoševičius and Herbert Jaeger. *Reservoir computing approaches to recurrent neural network training.* Computer Science Review, vol. 3, no. 3, pages 127–149, August 2009.

[Lynch *et al.* 1977] Gary S Lynch, Thomas Dunwiddie and Valentin Gribkoff. *Heterosynaptic depression: a postsynaptic correlate of long-term potentiation.* Nature, vol. 266, no. 5604, pages 737–739, April 1977.

[Ma & Pouget 2008] Wei Ji Ma and Alexandre Pouget. *Linking neurons to behavior in multisensory perception: A computational review.* Brain Research, vol. 1242, pages 4–12, November 2008.

[Ma *et al.* 2006] Wei Ji Ma, Jeffrey M Beck, Peter E Latham and Alexandre Pouget. *Bayesian inference with probabilistic population codes.* Nature Neuroscience, vol. 9, no. 11, pages 1432–1438, October 2006.

[Ma *et al.* 2008] Wei Ji Ma, Jeffrey M Beck and Alexandre Pouget. *Spiking networks for Bayesian inference and choice.* Current Opinion in Neurobiology, vol. 18, no. 2, pages 217–222, April 2008.

[Maass *et al.* 2002] Wolfgang Maass, Thomas Natschläger and Henry Markram. *Real-time computing without stable states: a new framework for neural computation based on perturbations.* Neural Computation, vol. 14, no. 11, pages 2531–60, November 2002.

[Maass *et al.* 2003] Wolfgang Maass, Thomas Natschläger and Henry Markram. *A model for real-time computation in generic neural microcircuits.* In Advances in Neural Information Processing Systems 15 (NIPS 2002), pages 213– 220, Vancouver, Canada, 2003. MIT Press, Cambidge, MA.

[Maass *et al.* 2006] Wolfgang Maass, Prashant Joshi and Eduardo D Sontag. *Principles of real-time computing with feedback applied to cortical microcircuit models.* In Advances in Neural Information Processing Systems 18 (NIPS 2005), pages 835–842, Vancouver, Canada, 2006. MIT Press, Cambidge, MA.

[MacKay 2003] David J C MacKay. Information Theory, Inference and Learning Algorithms. Cambridge University Press, 2003.

[Maffei *et al.* 2006] Arianna Maffei, Kiran Nataraj, Sacha B Nelson and Gina G Turrigiano. *Potentiation of cortical inhibition by visual deprivation.* Nature, vol. 443, no. 7107, pages 81–4, September 2006.

[Maffei 2011] Arianna Maffei. *The many forms and functions of long term plasticity at GABAergic synapses.* Neural Plasticity, vol. 2011, page 254724, January 2011.

[Magosso *et al.* 2008] Elisa Magosso, Cristiano Cuppini, Andrea Serino, Giuseppe Di Pellegrino and Mauro Ursino. *A theoretical study of multisensory integration in the superior colliculus by a neural network model.* Neural Networks, vol. 21, no. 6, pages 817–829, August 2008.

[Mamassian & Goutcher 2001] Pascal Mamassian and Ross Goutcher. *Prior knowledge on the illumination position.* Cognition, vol. 81, no. 1, pages B1–9, August 2001.

[Mamassian & Landy 1998] Pascal Mamassian and Michael S Landy. *Observer biases in the 3D interpretation of line drawings.* Vision Research, vol. 38, no. 18, pages 2817–2832, September 1998.

[Mamassian & Landy 2001] Pascal Mamassian and Michael S Landy. *Interaction of visual prior constraints.* Vision Research, vol. 41, no. 20, pages 2653–2668, September 2001.

[Markram *et al.* 1995] Henry Markram, P Johannes Helm and Bert Sakmann. *Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons.* Journal of Physiology, vol. 485, pages 1–20, May 1995.

[Markram *et al.* 1997] Henry Markram, Joachim Lübke, Michael Frotscher and Bert Sakmann. *Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs.* Science, vol. 275, no. 5297, pages 213–215, January 1997.

[Markram *et al.* 1998] Henry Markram, Yun Wang and Misha Tsodyks. *Differential signaling via the same axon of neocortical pyramidal neurons.* Proceedings of the National Academy of Sciences of the United States of America, vol. 95, no. 9, pages 5323–5328, April 1998.

[Markram *et al.* 2004] Henry Markram, Maria Toledo-Rodriguez, Yun Wang, Anirudh Gupta, Gilad Silberberg and Caizhi Wu. *Interneurons of the neocortical inhibitory system.* Nature Reviews Neuroscience, vol. 5, no. 10, pages 793–807, October 2004.

[Marr 1982] David Marr. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. MIT Press, Cambridge, UK, 1982.

[Matsumoto & Tanaka 2004] Kenji Matsumoto and Keiji Tanaka. *The role of the medial prefrontal cortex in achieving goals.* Current Opinion in Neurobiology, vol. 14, no. 2, pages 178–85, April 2004.

[McClure *et al.* 2003] Samuel M McClure, Gregory S Berns and P Read Montague. *Temporal prediction errors in a passive learning task activate human striatum.* Neuron, vol. 38, no. 2, pages 339–46, April 2003.

[Meltzoff & Borton 1979] AN Meltzoff and RW Borton. *Intermodal matching by human neonates.* Nature, vol. 282, no. 5737, pages 403–404, November 1979.

[Meredith & Stein 1983] M Alex Meredith and Barry E Stein. *Interactions among converging sensory inputs in the superior colliculus.* Science, vol. 221, no. 4608, pages 389–391, July 1983.

[Meredith & Stein 1986] M Alex Meredith and Barry E Stein. *Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration.* Journal of Neurophysiology, vol. 56, no. 3, pages 640–62, September 1986.

[Meredith *et al.* 1987] M Alex Meredith, J W Nemitz and Barry E Stein. *Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors.* Journal of Neuroscience, vol. 7, no. 10, pages 3215–3229, October 1987.

[Michel & Jacobs 2007] Melchi M Michel and Robert A Jacobs. *Parameter learning but not structure learning: a Bayesian network model of constraints on early perceptual learning.* Journal of Vision, vol. 7, no. 1, page 4, 2007.

[Miller 1982] Jeff Miller. *Divided attention: evidence for coactivation with redundant signals.* Cognitive Psychology, vol. 14, no. 2, pages 247–279, April 1982.

[Molina-Luna *et al.* 2009] Katiuska Molina-Luna, Ana Pekanovic, Sebastian Röhrich, Benjamin Hertler, Maximilian Schubring-Giese, Mengia-Seraina Rioult-Pedotti and Andreas R Luft. *Dopamine in motor cortex is necessary for skill learning and synaptic plasticity.* PLoS ONE, vol. 4, no. 9, page e7082, January 2009.

[Monaci *et al.* 2009] Gianluca Monaci, Pierre Vandergheynst and Friedrich T Sommer. *Learning bimodal structure in audio-visual data.* IEEE Transactions on Neural Networks, vol. 20, no. 12, pages 1898–910, December 2009.

[Montague *et al.* 1996] P Read Montague, Peter Dayan and Terrence J Sejnowski. *A framework for mesencephalic dopamine systems based on predictive Hebbian learning.* Journal of neuroscience, vol. 16, no. 5, pages 1936–47, March 1996.

[Moreno-Bote *et al.* 2011] Rubén Moreno-Bote, David C Knill and Alexandre Pouget. *Bayesian sampling in visual perception.* Proceedings of the National Academy of Sciences of the United States of America, vol. 108, no. 30, pages 12491–12496, July 2011.

[Morgenstern *et al.* 2011] Yaniv Morgenstern, Richard F Murray and Laurence R Harris. *The human visual system's assumption that light comes from above is weak.* Proceedings of the National Academy of Sciences, vol. 108, no. 30, pages 12551–12553, July 2011.

[Morris *et al.* 2006] Genela Morris, Alon Nevet, David Arkadir, Eilon Vaadia and Hagai Bergman. *Midbrain dopamine neurons encode decisions for future action.* Nature Neuroscience, vol. 9, no. 8, pages 1057–63, August 2006.

[Morrongiello *et al.* 1998a] Barbara A Morrongiello, Kimberley D Fenwick and Graham Chance. *Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events.* Infant Behavior and Development, vol. 21, no. 4, pages 543–553, 1998.

[Morrongiello *et al.* 1998b] Barbara A Morrongiello, Kimberley D Fenwick and Tanya Nutley. *Developmental changes in associations between auditory-visual events.* Infant Behavior and Development, vol. 21, no. 4, pages 613–626, 1998.

[Mozer *et al.* 2008] Michael C Mozer, Harold Pashler and Hadjar Homaei. *Optimal Predictions in Everyday Cognition: The Wisdom of Individuals or Crowds?* Cognitive Science, vol. 32, no. 7, pages 1133–1147, 2008.

[Murata *et al.* 1965] K Murata, H Cramer and Paul Bach-y Rita. *Neuronal convergence of noxious, acoustic, and visual stimuli in the visual cortex of the cat.* Journal of Neurophysiology, vol. 28, no. 6, pages 1223–39, November 1965.

[Mysore & Quartz 2005] S.P. Mysore and S.R. Quartz. *Modeling structural plasticity in the barn owl auditory localization system with a spike-time dependent Hebbian learning rule.* In International Joint Conference on Neural Networks (IJCNN 2005), volume 5, pages 2766–2771 vol. 5. IEEE, December 2005.

[Nakadai *et al.* 2006] Kazuhiro Nakadai, Hirofumi Nakajima, Masamitsu Murase, Hiroshi Okuno, Yuji Hasegawa and Hiroshi Tsujino. *Real-Time Tracking of Multiple Sound Sources by Integration of In-Room and Robot-Embedded Microphone Arrays.* In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006), pages 852–859, Beijing, China, October 2006. IEEE.

[Nakahara *et al.* 2004] Hiroyuki Nakahara, Hideaki Itoh, Reiko Kawagoe, Yoriko Takikawa and Okihide Hikosaka. *Dopamine neurons can represent context-dependent prediction error.* Neuron, vol. 41, no. 2, pages 269–80, January 2004.

[Nardini *et al.* 2008] Marko Nardini, Peter Jones, Rachael Bedford and Oliver Braddick. *Development of cue integration in human navigation.* Current Biology, vol. 18, no. 9, pages 689–693, May 2008.

[Nardini *et al.* 2010] Marko Nardini, Rachael Bedford and Denis Mareschal. *Fusion of visual cues is not mandatory in children.* Proceedings of the National Academy of Sciences of the United States of America, vol. 107, no. 39, pages 17041–17046, September 2010.

[Needham 1999] Amy Needham. *The role of shape in 4-month-old infants' object segregation.* Infant Behavior and Development, vol. 22, no. 2, pages 161–178, 1999.

[Neil *et al.* 2006] Patricia A. Neil, Christine Chee-Ruiter, Christian Scheier, David J. Lewkowicz and Shinsuke Shimojo. *Development of multisensory spatial integration and perception in humans.* Developmental Science, vol. 9, no. 5, pages 454–464, September 2006.

[Nelson & Turrigiano 2008] Sacha B Nelson and Gina G Turrigiano. *Strength through diversity.* Neuron, vol. 60, no. 3, pages 477–82, November 2008.

[Nickerson 1973] Raymond S Nickerson. *Intersensory facilitation of reaction time: Energy summation or preparation enhancement?* Psychological Review, vol. 80, no. 6, pages 489–509, 1973.

[Norton & Ventura 2006] David Norton and Dan Ventura. *Preparing More Effective Liquid State Machines Using Hebbian Learning.* In International Joint Conference on Neural Network (IJCNN 2006), pages 4243–4248. IEEE, 2006.

[O'Doherty *et al.* 2004] John O'Doherty, Peter Dayan, Johannes Schultz, Ralf Deichmann, Karl Friston and Raymond J Dolan. *Dissociable roles of ventral and dorsal striatum in instrumental conditioning.* Science, vol. 304, no. 5669, pages 452–4, April 2004.

[Ohshiro *et al.* 2011] Tomokazu Ohshiro, Dora E Angelaki and Gregory C Deangelis. *A normalization model of multisensory integration.* Nature Neuroscience, vol. 14, no. 6, pages 775–82, June 2011.

[Olshausen & Field 2004] Bruno A Olshausen and David J Field. *Sparse coding of sensory inputs.* Current Opinion in Neurobiology, vol. 14, no. 4, pages 481–7, August 2004.

[Oruç *et al.* 2003] Ipek Oruç, Laurence T Maloney and Michael S Landy. *Weighted linear cue combination with possibly correlated error.* Vision Research, vol. 43, no. 23, pages 2451–2468, October 2003.

[Patton & Anastasio 2003] Paul E Patton and Thomas J Anastasio. *Modeling cross-modal enhancement and modality-specific suppression in multisensory neurons.* Neural Computation, vol. 15, no. 4, pages 783–810, April 2003.

[Paugam-Moisy *et al.* 2008] Hélène Paugam-Moisy, Régis Martinez and Samy Bengio. *Delay learning and polychronization for reservoir computing.* Neurocomputing, vol. 71, no. 7-9, pages 1143–1158, March 2008.

[Pavlov 1926] Ivan Petrovich Pavlov. Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex. Oxford University Press, New York, NY, USA, 1926.

[Pawlak & Kerr 2008] Verena Pawlak and Jason N D Kerr. *Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity.* Journal of Neuroscience, vol. 28, no. 10, pages 2435–46, March 2008.

[Pawlak *et al.* 2010] Verena Pawlak, Jeffery R Wickens, Alfredo Kirkwood and Jason N D Kerr. *Timing is not everything: neuromodulation opens the STDP gate.* Frontiers in Synaptic Neuroscience, vol. 2, no. 146, 2010.

[Paz *et al.* 2009] Jeanne T Paz, Séverine Mahon, Pascale Tiret, Stéphane Genet, Bruno Delord and Stéphane Charpier. *Multiple forms of activity-dependent intrinsic plasticity in layer V cortical neurones in vivo.* Journal of Physiology, vol. 587, no. Pt 13, pages 3189–205, July 2009.

[Perez-Orive *et al.* 2002] Javier Perez-Orive, Ofer Mazor, Glenn C Turner, Stijn Cassenaer, Rachel I Wilson and Gilles Laurent. *Oscillations and sparsening of odor representations in the mushroom body.* Science, vol. 297, no. 5580, pages 359–65, July 2002.

[Perfors *et al.* 2011] Amy Perfors, Joshua B Tenenbaum, Thomas L Griffiths and Fei Xu. *A tutorial introduction to Bayesian models of cognitive development.* Cognition, vol. 120, no. 3, pages 302–321, January 2011.

[Pessiglione *et al.* 2006] Mathias Pessiglione, Ben Seymour, Guillaume Flandin, Raymond J Dolan and Chris D Frith. *Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans.* Nature, vol. 442, no. 7106, pages 1042–5, August 2006.

[Pessiglione *et al.* 2008] Mathias Pessiglione, Predrag Petrovic, Jean Daunizeau, Stefano Palminteri, Raymond J Dolan and Chris D Frith. *Subliminal instrumental conditioning demonstrated in the human brain.* Neuron, vol. 59, no. 4, pages 561–7, August 2008.

[Pfeiffer *et al.* 2010] Michael Pfeiffer, Bernhard Nessler, Rodney J. Douglas and Wolfgang Maass. *Reward-Modulated Hebbian Learning of Decision Making.* Neural Computation, vol. 22, no. 6, pages 1399–1444, June 2010.

[Phillips *et al.* 1976] Anthony G Phillips, David A Carter and Hans C Fibiger. *Dopaminergic substrates of intracranial self-stimulation in the caudate-putamen.* Brain Research, vol. 104, no. 2, pages 221–32, March 1976.

[Platt & Glimcher 1999] ML Platt and PW Glimcher. *Neural correlates of decision variables in parietal cortex.* Nature, vol. 400, no. 6741, pages 233–238, July 1999.

[Pons *et al.* 2009] Ferran Pons, David J Lewkowicz, Salvador Soto-Faraco and Núria Sebastián-Gallés. *Narrowing of intersensory speech perception in infancy.* Proceedings of the National Academy of Sciences of the United States of America, vol. 106, no. 26, pages 10598–602, June 2009.

[Pratt & Aizenman 2007] Kara G Pratt and Carlos D Aizenman. *Homeostatic regulation of intrinsic excitability and synaptic transmission in a developing visual circuit.* Journal of Neuroscience, vol. 27, no. 31, pages 8268–77, August 2007.

[Putzar *et al.* 2007] Lisa Putzar, Ines Goerendt, Kathrin Lange, Frank Rösler and Brigitte Röder. *Early visual deprivation impairs multisensory interactions in humans.* Nature Neuroscience, vol. 10, no. 10, pages 1243–1245, October 2007.

[Raab 1962] D H Raab. *Statistical facilitation of simple reaction times.* Transactions of the New York Academy of Sciences, vol. 24, pages 574–90, March 1962.

[Rach *et al.* 2011] Stefan Rach, Adele Diederich and Hans Colonius. *On quantifying multisensory interaction effects in reaction time and detection rate.* Psychological Research, vol. 75, no. 2, pages 77–94, March 2011.

[Rainer *et al.* 1998] Gregor Rainer, Wael F Asaad and Earl K Miller. *Selective representation of relevant information by neurons in the primate prefrontal cortex.* Nature, vol. 393, no. 6685, pages 577–9, June 1998.

[Rangel & Hare 2010] Antonio Rangel and Todd Hare. *Neural computations associated with goal-directed choice.* Current Opinion in Neurobiology, vol. 20, no. 2, pages 262–270, March 2010.

[Rao & Ballard 1999] Rajesh P N Rao and Dana H Ballard. *Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects.* Nature Neuroscience, vol. 2, no. 1, pages 79–87, January 1999.

[Rao 1973] C Radhakrishna Rao. Linear Statistical Inference and its Applications. Wiley-Interscience, New York, 2nd édition, 1973.

[Rao 2004] Rajesh P Rao. *Bayesian computation in recurrent neural circuits.* Neural Computation, vol. 16, no. 1, pages 1–38, January 2004.

[Rao 2010] Rajesh P Rao. *Decision Making under Uncertainty: A Neural Model based on Partially Observable Markov Decision Processes.* Frontiers in Computational Neuroscience, vol. 4, no. 146, 2010.

[Recanzone 1998] Gregg H Recanzone. *Rapidly induced auditory plasticity: the ventriloquism aftereffect.* Proceedings of the National Academy of Sciences of the United States of America, vol. 95, no. 3, pages 869–875, February 1998.

[Reuschel *et al.* 2010] Johanna Reuschel, Knut Drewing, Denise Henriques, Frank Rösler and Katja Fiehler. *Optimal integration of visual and proprioceptive movement information for the perception of trajectory geometry.* Experimental Brain Research, vol. 201, no. 4, pages 853–862, April 2010.

[Reynolds & Wickens 2002] John N J Reynolds and Jeffery R Wickens. *Dopamine-dependent plasticity of corticostriatal synapses.* Neural Networks, vol. 15, no. 4-6, pages 507–21, 2002.

[Reynolds *et al.* 2001] John N J Reynolds, Brian I Hyland and Jeffery R Wickens. *A cellular mechanism of reward-related learning.* Nature, vol. 413, no. 6851, pages 67–70, September 2001.

[Rodemann *et al.* 2008] Tobias Rodemann, Gökhan Ince, Frank Joublin and Christian Goerick. *Using Binaural and Spectral Cues for Azimuth and Elevation Localization.* In IEEE/RSJ International Conference on Intelligent Robot and Systems (IROS 2008), pages 2185–2190. IEEE, 2008.

[Rodemann 2010] Tobias Rodemann. *A study on distance estimation in binaural sound localization.* In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), Taipei, 2010. IEEE.

[Roelfsema & van Ooyen 2005] Pieter R Roelfsema and Arjen van Ooyen. *Attention-gated reinforcement learning of internal representations for classification.* Neural Computation, vol. 17, no. 10, pages 2176–214, October 2005.

[Roesch *et al.* 2007] Matthew R Roesch, Donna J Calu and Geoffrey Schoenbaum. *Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards.* Nature Neuroscience, vol. 10, no. 12, pages 1615–24, December 2007.

[Rojas 2010] Randall Rojas Rojas. *Explaining Human Causal Learning using a Dynamic Probabilistic Model.* PhD thesis, University of California, Los Angeles, 2010.

[Ross *et al.* 2011] Lars A Ross, Sophie Molholm, Daniella Blanco, Manuel Gomez-Ramirez, Dave Saint-Amour and John J Foxe. *The development of multisensory speech perception continues into the late childhood years.* European Journal of Neuroscience, vol. 33, no. 12, pages 2329–2337, 2011.

[Rothkopf *et al.* 2009] Constantin A Rothkopf, Thomas H Weisswange and Jochen Triesch. *Learning independent causes in natural images explains the spacevariant oblique effect.* In 8th International Conference on Development and Learning, pages 1–6, Shanghai, China, 2009. IEEE.

[Rothkopf *et al.* 2010] Constantin A Rothkopf, Thomas H Weisswange and Jochen Triesch. *Computational modeling of multisensory object perception.* In MJ Naumer and J Kaiser, editeurs, Multisensory Object Perception in the Primate Brain, chapitre 3, pages 21–53. Springer, New York, 2010.

[Rowland *et al.* 2007a] Benjamin A Rowland, Stephan Quessy, Terrence R Stanford and Barry E Stein. *Multisensory Integration Shortens Physiological Response Latencies.* Journal of Neuroscience, vol. 27, no. 22, pages 5879–5884, May 2007.

[Rowland *et al.* 2007b] Benjamin A Rowland, Terrence R Stanford and Barry E Stein. *A Bayesian model unifies multisensory spatial localization with the physiological properties of the superior colliculus.* Experimental Brain Research, vol. 180, no. 1, pages 153–161, June 2007.

[Rowland *et al.* 2007c] Benjamin A Rowland, Terrence R Stanford and Barry E Stein. *A model of the neural mechanisms underlying multisensory integration in the superior colliculus.* Perception, vol. 36, no. 10, pages 1431–1443, 2007.

[Rozeboom 1958] William W Rozeboom. *What is learned? An empirical enigma.* Psychological Review, vol. 65, no. 1, pages 22–33, January 1958.

[Rubinstein 1959] Irvin Rubinstein. *Some factors in probability matching.* Journal of Experimental Psychology, vol. 57, no. 6, pages 413–6, June 1959.

[Rucci *et al.* 1997] Michele Rucci, Giulio Tononi and Gerald M Edelman. *Registration of Neural Maps through Value-Dependent Learning: Modeling the Alignment of Auditory and Visual Maps in the Barn Owl's Optic Tectum.* Journal of Neuroscience, vol. 17, no. 1, pages 334–352, January 1997.

[Rucci *et al.* 2000] M Rucci, J Wray and GM Edelman. *Robust localization of auditory and visual targets in a robotic barn owl.* Robotics and Autonomous Systems, vol. 30, no. 1-2, pages 181–193, January 2000.

[Rumelhart *et al.* 1986a] David E Rumelhart, Geoffrey E Hinton and Ronald J Williams. *Learning internal representations by error propagation.* In David E Rumelhart and James L McClelland, editeurs, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, pages 318–362. MIT Press, 1st édition, 1986.

[Rumelhart *et al.* 1986b] David E Rumelhart, Geoffrey E Hinton and Ronald J Williams. *Learning representations by back-propagating errors.* Nature, vol. 323, no. 6088, pages 533–536, October 1986.

[Rushworth & Behrens 2008] Matthew F S Rushworth and Timothy E J Behrens. *Choice, uncertainty and value in prefrontal and cingulate cortex.* Nature Neuroscience, vol. 11, no. 4, pages 389–97, April 2008.

[Sahani & Dayan 2003] Maneesh Sahani and Peter Dayan. *Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity.* Neural Computation, vol. 15, no. 10, pages 2255–2279, October 2003.

[Samejima & Doya 2007] Kazuyuki Samejima and Kenji Doya. *Multiple representations of belief states and action values in corticobasal ganglia loops.* Annals of the New York Academy of Sciences, vol. 1104, pages 213–28, May 2007.

[Samejima *et al.* 2005] Kazuyuki Samejima, Yasumasa Ueda, Kenji Doya and Minoru Kimura. *Representation of action-specific reward values in the striatum.* Science, vol. 310, no. 5752, pages 1337–40, November 2005.

[Sasaki *et al.* 2006] Yoko Sasaki, Satoshi Kagami and Hiroshi Mizoguchi. *Multiple Sound Source Mapping for a Mobile Robot by Self-motion Triangulation.* In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006), pages 380–385, Beijing, China, October 2006. IEEE.

[Sato *et al.* 2007] Yoshiyuki Sato, Taro Toyoizumi and Kazuyuki Aihara. *Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli.* Neural Computation, vol. 19, no. 12, pages 3335–3355, December 2007.

[Savazzi & Marzi 2002] Silvia Savazzi and Carlo A Marzi. *Speeding up reaction time with invisible stimuli.* Current Biology, vol. 12, no. 5, pages 403–407, March 2002.

[Savin & Triesch 2010] Cristina Savin and Jochen Triesch. *Structural plasticity improves stimulus encoding in a working memory model.* In Computational and Systems Neuroscience (COSYNE 2010). Frontiers in Systems Neuroscience, 2010.

[Savin *et al.* 2010] Cristina Savin, Prashant Joshi and Jochen Triesch. *Independent component analysis in spiking neurons.* PLoS Computational Biology, vol. 6, no. 4, page e1000757, January 2010.

[Savin 2010] Cristina Savin. *Homeostatic plasticity - computational and clinical implications.* PhD thesis, Goethe University, Frankfurt, Germany, 2010.

[Scharstein *et al.* 2002] D Scharstein, R Szeliski and R Zabih. *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms.* Workshop on Stereo and Multi-Baseline Vision (SMBV 2001), vol. 47, pages 131–140, 2002.

[Schauer & Gross 2003] C. Schauer and H.-M. Gross. *A Computational Model of Early Auditory-Visual Integration.* In Pattern Recognition, pages 362–369. 2003.

[Schiller & Steil 2005] Ulf D Schiller and Jochen J Steil. *Analyzing the weight dynamics of recurrent learning algorithms.* Neurocomputing, vol. 63, pages 5–23, January 2005.

[Schliebs *et al.* 2011] Stefan Schliebs, Mohemmed Ammar and Nikola Kasabov. *Are probabilistic spiking neural networks suitable for reservoir computing?* In International Joint Conference on Neural Networks (IJCNN 2011), San Jose, California, USA, 2011.

[Schmidhuber *et al.* 2007] Jürgen Schmidhuber, Daan Wierstra, Matteo Gagliolo and Faustino Gomez. *Training recurrent networks by Evolino.* Neural Computation, vol. 19, no. 3, pages 757–79, March 2007.

[Schönberg *et al.* 2007] Tom Schönberg, Nathaniel D Daw, Daphna Joel and John P O'Doherty. *Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making.* Journal of Neuroscience, vol. 27, no. 47, pages 12860–7, November 2007.

[Schorr *et al.* 2005] Efrat A Schorr, Nathan A Fox, Virginie van Wassenhove and Eric I Knudsen. *Auditory-visual fusion in speech perception in children with cochlear implants.* Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 51, pages 18748–50, December 2005.

[Schrauwen *et al.* 2008] Benjamin Schrauwen, Marion Wardermann, David Verstraeten, Jochen J Steil and Dirk Stroobandt. *Improving reservoirs using intrinsic plasticity.* Neurocomputing, vol. 71, no. 7-9, pages 1159–1171, March 2008.

[Schultz *et al.* 1997] Wolfram Schultz, Peter Dayan and P Read Montague. *A neural substrate of prediction and reward.* Science, vol. 275, no. 5306, pages 1593–9, 1997.

[Schultz 1986] Wolfram Schultz. *Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey.* Journal of Neurophysiology, vol. 56, no. 5, pages 1439–61, November 1986.

[Schultz 2000] Wolfram Schultz. *Multiple reward signals in the brain.* Nature Reviews Neuroscience, vol. 1, no. 3, pages 199–207, 2000.

[Sealfon & Olanow 2000] Stuart C Sealfon and Warren Olanow. *Dopamine receptors: from structure to behavior.* Trends in Neurosciences, vol. 23, no. Supplement 1, pages S34–40, October 2000.

[Seitz *et al.* 2007] Aaron R Seitz, Robyn S Kim, Virginie van Wassenhove and Ladan Shams. *Simultaneous and independent acquisition of multisensory and unisensory associations.* Perception, vol. 36, no. 10, pages 1445–1453, 2007.

[Seydell *et al.* 2010] Anna Seydell, David C Knill and Julia Trommershäuser. *Adapting internal statistical models for interpreting visual cues to depth.* Journal of Vision, vol. 10, no. 4, pages 1.1–27, January 2010.

[Shah & Barto 2009] Ashvin Shah and Andrew G Barto. *Effect on movement selection of an evolving sensory representation: a multiple controller model of skill acquisition.* Brain Research, vol. 1299, pages 55–73, November 2009.

[Shams & Beierholm 2010] Ladan Shams and Ulrik R Beierholm. *Causal inference in perception.* Trends in Cognitive Sciences, vol. 14, no. 9, pages 425–432, August 2010.

[Shen *et al.* 2008] Weixing Shen, Marc Flajolet, Paul Greengard and D James Surmeier. *Dichotomous dopaminergic control of striatal synaptic plasticity.* Science, vol. 321, no. 5890, pages 848–51, August 2008.

[Shepherd *et al.* 2006] Jason D Shepherd, Gavin Rumbaugh, Jing Wu, Shoaib Chowdhury, Niels Plath, Dietmar Kuhl, Richard L Huganir and Paul F Worley. *Arc/Arg3.1 mediates homeostatic synaptic scaling of AMPA receptors.* Neuron, vol. 52, no. 3, pages 475–84, November 2006.

[Shi & Griffiths 2010] Lei Shi and Thomas L Griffiths. *Neural Implementation of Hierarchical Bayesian Inference by Importance Sampling.* In Y Bengio, D Schuurmans, J Lafferty, C K I Williams and A Culotta, editeurs, Advances in Neural Information Processing Systems 22 (NIPS 2009), pages 1669–1677, Vancouver, Canada, 2010.

[Singer & Gray 1995] Wolf Singer and Charles M Gray. *Visual feature integration and the temporal correlation hypothesis.* Annual Review of Neuroscience, vol. 18, pages 555–86, January 1995.

[Singh *et al.* 2000] Satinder Singh, Tommi Jaakkola, Michael L Littman and Csaba Szepesvári. *Convergence Results for Single-Step On-Policy Reinforcement-Learning Algorithms.* Machine Learning, vol. 38, no. 3, pages 287–308, March 2000.

[Sinha & Adelson 1993] Pawan Sinha and Edward H Adelson. *Recovering reflectance and illumination in a world of painted polyhedra.* In 4th International Conference on Computer Vision (ICCV 1993), pages 156–163, Berlin , Germany, 1993. IEEE Computer Society Press.

[Slee & Young 2011] Sean J Slee and Eric D Young. *Information conveyed by inferior colliculus neurons about stimuli with aligned and misaligned sound localization cues.* Journal of Neurophysiology, vol. 106, no. 2, pages 974–985, August 2011.

[Sloman 1996] Steven A Sloman. *The Empirical Case for Two Systems of Reasoning.* Psychological Bulletin, vol. 119, no. 1, pages 3–22, 1996.

[Slutsky & Recanzone 2001] Daniel A Slutsky and Gregg H Recanzone. *Temporal and spatial dependency of the ventriloquism effect.* Neuroreport, vol. 12, no. 1, pages 7–10, January 2001.

[Smith 2001] A. Mark Smith. Alhacen's Theory of Visual Perception (First Three Books of Alhacen's De Aspectibus). American Philosophical Society, 2001.

[Sobel *et al.* 2004] David M Sobel, Joshua B Tenenbaum and Alison Gopnik. *Children's Causal Inferences from Indirect Evidence: Backwards Blocking and Bayesian Reasoning in Preschoolers.* Cognitive Science, vol. 28, no. 3, pages 303–333, 2004.

[Soltani & Wang 2010] Alireza Soltani and Xiao-Jing Wang. *Synaptic computation underlying probabilistic inference.* Nature Neuroscience, vol. 13, no. 1, pages 112–119, January 2010.

[Song *et al.* 2000] Sen Song, Kenneth D Miller and Larry F Abbott. *Competitive Hebbian learning through spike-timing-dependent synaptic plasticity.* Nature Neuroscience, vol. 3, no. 9, pages 919–26, September 2000.

[Sparks & Mays 1990] David L Sparks and Lawrence E Mays. *Signal transformations required for the generation of saccadic eye movements.* Annual Review of Neuroscience, vol. 13, pages 309–336, 1990.

[Spelke 1976] Elizabeth S Spelke. *Infants' intermodal perception of events.* Cognitive Psychology, vol. 8, no. 4, pages 553–560, October 1976.

[Sprekeler *et al.* 2011] Henning Sprekeler, Tim Vogels, Claudia Clopath and Wulfram Gerstner. *Balancing excitation and inhibition: A theoretical analysis of inhibitory synaptic plasticity.* In Computational and Systems Neuroscience (COSYNE 2011), Salt Lake City, USA, 2011.

[Stanford *et al.* 2005] Terrence R Stanford, Stephan Quessy and Barry E Stein. *Evaluating the operations underlying multisensory integration in the cat superior colliculus.* Journal of Neuroscience, vol. 25, no. 28, pages 6499–6508, July 2005.

[Steil 2004] Jochen J Steil. *Backpropagation-decorrelation: online recurrent learning with O(N) complexity.* In International Joint Conference on Neural Networks (IJCNN 2004), volume 2, pages 843–848. IEEE, 2004.

[Steil 2007] Jochen J Steil. *Online reservoir adaptation by intrinsic plasticity for backpropagation-decorrelation and echo state learning.* Neural Networks, vol. 20, no. 3, pages 353–64, April 2007.

[Stein *et al.* 1989] Barry E Stein, M Alex Meredith, W Scott Huneycutt and Lawrence McDade. *Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli.* Journal of Cognitive Neuroscience, vol. 1, no. 1, pages 12–24, December 1989.

[Stein *et al.* 1993] Barry E Stein, M Alex Meredith and Mark T Wallace. *The visually responsive neuron and beyond: multisensory integration in cat and monkey.* Progress in Brain Research, vol. 95, pages 79–90, 1993.

[Stemmler & Koch 1999] Martin Stemmler and Christof Koch. *How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate.* Nature Neuroscience, vol. 2, no. 6, pages 521–7, June 1999.

[Stocker & Simoncelli 2006] Alan A Stocker and Eero P Simoncelli. *Noise characteristics and prior expectations in human visual speed perception.* Nature Neuroscience, vol. 9, no. 4, pages 578–585, April 2006.

[Strelnikov *et al.* 2011] Kuzma Strelnikov, Maxime Rosito and Pascal Barone. *Effect of Audiovisual Training on Monaural Spatial Hearing in Horizontal Plane.* PLoS ONE, vol. 6, no. 3, page e18344, March 2011.

[Summerfield *et al.* 2011] Christopher Summerfield, Timothy E Behrens and Etienne Koechlin. *Perceptual Classification in a Rapidly Changing Environment.* Neuron, vol. 71, no. 4, pages 725–736, 2011.

[Surmeier *et al.* 2007] D James Surmeier, Jun Ding, Michelle Day, Zhongfeng Wang and Weixing Shen. *D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons.* Trends in Neurosciences, vol. 30, no. 5, pages 228–35, May 2007.

[Sutton & Barto 1998] Richard S Sutton and Andrew G Barto. Reinforcement Learning: An Introduction. MIT Press, Cambridge, UK, 1998.

[Sutton 1988] Richard S Sutton. *Learning to predict by the methods of temporal differences.* Machine Learning, vol. 3, no. 1, pages 9–44, August 1988.

[Takikawa *et al.* 2002] Yoriko Takikawa, Reiko Kawagoe, Hideaki Itoh, Hiroyuki Nakahara and Okihide Hikosaka. *Modulation of saccadic eye movements by predicted reward outcome.* Experimental Brain Research, vol. 142, no. 2, pages 284–91, 2002.

[Talsma *et al.* 2010] Durk Talsma, Daniel Senkowski, Salvador Soto-Faraco and Marty G Woldorff. *The multifaceted interplay between attention and multisensory integration.* Trends in cognitive sciences, vol. 14, no. 9, pages 400–410, August 2010.

[Tao *et al.* 2000] Hui-zhong W Tao, Li I Zhang, Guo-qiang Bi and Mu-ming Poo. *Selective presynaptic propagation of long-term potentiation in defined neural networks.* Journal of Neuroscience, vol. 20, no. 9, pages 3233–43, May 2000.

[Taylor 1962] M M Taylor. *Figural after-effects: a psychophysical theory of the displacement effect.* Canadian Journal of Psychology, vol. 16, pages 247–77, December 1962.

[Teder-Sälejärvi *et al.* 2005] Wolfgang A Teder-Sälejärvi, Francesco Di Russo, J J McDonald and Steven A Hillyard. *Effects of spatial congruity on audio-visual multimodal integration.* Journal of Cognitive Neuroscience, vol. 17, no. 9, pages 1396–1409, September 2005.

[Tenenbaum *et al.* 2011] Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths and Noah D Goodman. *How to Grow a Mind: Statistics, Structure, and Abstraction.* Science, vol. 331, no. 6022, pages 1279–1285, March 2011.

[Thomas 1941] G J Thomas. *Experimental study of the influence of vision on sound localization.* Journal of Experimental Psychology, vol. 28, no. 2, pages 163–177, 1941.

[Thorndike 1911] Edward Lee Thorndike. Animal intelligence: experimental studies. 1911.

[Todd 1912] J W Todd. *Reaction to multiple stimuli.* Archives of Psychology, vol. 25, pages 1–65, 1912.

[Tolman & Gleitman 1949] Edward C Tolman and Henry Gleitman. *Studies in learning and motivation: I. Equal reinforcements in both end-boxes, followed by shock in one end-box.* Journal of Experimental Psychology, vol. 39, no. 6, pages 810–9, December 1949.

[Treisman & Gelade 1980] Anne M Treisman and Garry Gelade. *A feature-integration theory of attention.* Cognitive Psychology, vol. 12, no. 1, pages 97–136, January 1980.

[Triesch & Eckes 1998] Jochen Triesch and Christian Eckes. *Object Recognition with Multiple Feature Types.* In 8th International Conference on Artificial Neural Networks (ICANN'98), pages 233–238, Skövde, Sweden, 1998.

[Triesch & von der Malsburg 2001] Jochen Triesch and Christoph von der Malsburg. *Democratic integration: self-organized integration of adaptive cues.* Neural Computation, vol. 13, no. 9, pages 2049–2074, September 2001.

[Triesch *et al.* 2002] Jochen Triesch, Dana H Ballard and Robert A Jacobs. *Fast temporal dynamics of visual cue integration.* Perception, vol. 31, no. 4, pages 421–434, 2002.

[Triesch *et al.* 2010] Jochen Triesch, Constantin A Rothkopf and Thomas H Weisswange. *Coordination in Sensory Integration.* In C Von Der Malsburg, William A Phillips and Wolf Singer, editeurs, Dynamic Coordination in the Brain, chapitre 15, pages 229–234. MIT Press, Cambridge, 2010.

[Triesch 2005a] Jochen Triesch. *A Gradient Rule for the Plasticity of a Neuronâs Intrinsic Excitability.* In 15th International Conference on Artificial Neural Networks (ICANN 2005), pages 1–7, Warsaw, Poland, 2005.

[Triesch 2005b] Jochen Triesch. *Synergies between intrinsic and synaptic plasticity in individual model neurons.* In Advances in Neural Information Processing Systems 17 (NIPS 2004), pages 1417–1424, Vancouver, Canada, 2005. MIT Press, Cambidge, MA.

[Triesch 2007] Jochen Triesch. *Synergies between intrinsic and synaptic plasticity mechanisms.* Neural Computation, vol. 19, no. 4, pages 885–909, April 2007.

[Turrigiano & Nelson 2004] Gina G Turrigiano and Sacha B Nelson. *Homeostatic plasticity in the developing nervous system.* Nature Reviews Neuroscience, vol. 5, no. 2, pages 97–107, March 2004.

[Turrigiano *et al.* 1998] Gina G Turrigiano, Kenneth R Leslie, Niraj S Desai, Lana C Rutherford and Sacha B Nelson. *Activity-dependent scaling of quantal amplitude in neocortical neurons.* Nature, vol. 391, no. 6670, pages 892–6, February 1998.

[Turrigiano 2008] Gina G Turrigiano. *The self-tuning neuron: synaptic scaling of excitatory synapses.* Cell, vol. 135, no. 3, pages 422–35, October 2008.

[Turrigiano 2011] Gina Turrigiano. *Too Many Cooks? Intrinsic and Synaptic Homeostatic Mechanisms in Cortical Circuit Refinement.* Annual Review of Neuroscience, vol. 34, pages 89–103, January 2011.

[Urbanczik & Senn 2009] Robert Urbanczik and Walter Senn. *Reinforcement learning in populations of spiking neurons.* Nature Neuroscience, vol. 12, no. 3, pages 250–252, February 2009.

[Ursino *et al.* 2008] Mauro Ursino, Cristiano Cuppini, Elisa Magosso, Andrea Serino and Giuseppe di Pellegrino. *Multisensory integration in the superior colliculus: a neural network model.* Journal of Computational Neuroscience, vol. 26, no. 1, pages 55–73, May 2008.

[Usher & McClelland 2001] Marius Usher and James L McClelland. *The time course of perceptual choice: the leaky, competing accumulator model.* Psychological Review, vol. 108, no. 3, pages 550–592, July 2001.

[Valentin *et al.* 2007] Vivian V Valentin, Anthony Dickinson and John P O'Doherty. *Determining the neural substrates of goal-directed learning in the human brain.* Journal of Neuroscience, vol. 27, no. 15, pages 4019–26, April 2007.

[van Beers *et al.* 1996] Rob J van Beers, Anne C Sittig and Jan J Denier van der Gon. *How humans combine simultaneous proprioceptive and visual position information.* Experimental Brain Research, vol. 111, no. 2, pages 253–61, September 1996.

[van Beers *et al.* 1999] Rob J van Beers, Anne C Sittig and Jan J Denier van der Gon. *Integration of proprioceptive and visual position-information: An experimentally supported model.* Journal of Neurophysiology, vol. 81, no. 3, pages 1355–1364, March 1999.

[van Rossum *et al.* 2000] Mark C W van Rossum, Guo-qiang Bi and Gina G Turrigiano. *Stable Hebbian learning from spike timing-dependent plasticity.* Journal of Neuroscience, vol. 20, no. 23, pages 8812–21, December 2000.

[van Vreeswijk & Sompolinsky 1996] Carl van Vreeswijk and Haim Sompolinsky. *Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity.* Science, vol. 274, no. 5293, pages 1724–1726, December 1996.

[Vasilaki *et al.* 2009] Eleni Vasilaki, Nicolas Frémaux, Robert Urbanczik, Walter Senn and Wulfram Gerstner. *Spike-based reinforcement learning in continuous state and action space: when policy gradient methods fail.* PLoS Computational Biology, vol. 5, no. 12, page e1000586, December 2009.

[Verstraeten *et al.* 2007] David Verstraeten, Benjamin Schrauwen, Michiel D'Haene and Dirk Stroobandt. *An experimental unification of reservoir computing methods.* Neural Networks, vol. 20, no. 3, pages 391–403, April 2007.

[Vogels *et al.* 2011] Tim Vogels, Henning Sprekeler, Friedemann Zenke, Claudia Clopath and Wulfram Gerstner. *Inhibitory synaptic plasticity generates global and detailed balance of excitation and inhibition.* In Computational and Systems Neuroscience (COSYNE 2011), Salt Lake City, USA, 2011.

[von der Malsburg 1973] Christoph von der Malsburg. *Self-organization of orientation sensitive cells in the striate cortex.* Kybernetik, vol. 14, no. 2, pages 85–100, December 1973.

[von der Malsburg 1995] Christoph von der Malsburg. *Binding in models of perception and brain function.* Current Opinion in Neurobiology, vol. 5, no. 4, pages 520–6, August 1995.

[von Helmholtz 1867] Hermann von Helmholtz. Handbuch der physiologischen Optik. Leopold Voss, Leipzig, Germany, 1867.

[Vul & Rich 2010] Edward Vul and Anina N Rich. *Independent sampling of features enables conscious perception of bound objects.* Psychological Science, vol. 21, no. 8, pages 1168–75, August 2010.

[Vul *et al.* 2009] Edward Vul, N D Goodman, Thomas L Griffiths and Joshua B Tenenbaum. *One and Done? Optimal Decisions From Very Few Samples.* In N A Taatgen and H van Rijn, editeurs, 31st Annual Conference of the Cognitive Science Society, Austin, TX, 2009. Cognitive Science Society.

[Wallace & Stein 1997] Mark T Wallace and Barry E Stein. *Development of multisensory neurons and multisensory integration in cat superior colliculus.* Journal of Neuroscience, vol. 17, no. 7, pages 2429–2444, April 1997.

[Wallace & Stein 2000] Mark T Wallace and Barry E Stein. *Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences.* Journal of Neurophysiology, vol. 83, no. 6, pages 3578–3582, June 2000.

[Wallace & Stein 2001] Mark T Wallace and Barry E Stein. *Sensory and multisensory responses in the newborn monkey superior colliculus.* Journal of Neuroscience, vol. 21, no. 22, pages 8886–8894, November 2001.

[Wallace & Stein 2007] Mark T Wallace and Barry E Stein. *Early experience determines how the senses will interact.* Journal of Neurophysiology, vol. 97, no. 1, pages 921–926, January 2007.

[Wallace *et al.* 1992] Mark T Wallace, M Alex Meredith and Barry E Stein. *Integration of multiple sensory modalities in cat cortex.* Experimental Brain Research, vol. 91, no. 3, pages 484–488, 1992.

[Wallace *et al.* 1996] Mark T Wallace, Lee K Wilkinson and Barry E Stein. *Representation and integration of multiple sensory inputs in primate superior colliculus.* Journal of Neurophysiology, vol. 76, no. 2, pages 1246–1266, August 1996.

[Wallace *et al.* 2004a] Mark T Wallace, Thomas J Perrault, W David Hairston and Barry E Stein. *Visual experience is necessary for the development of multisensory integration.* Journal of Neuroscience, vol. 24, no. 43, pages 9580–9584, October 2004.

[Wallace *et al.* 2004b] Mark T Wallace, G E Roberson, W David Hairston, Barry E Stein, J William Vaughan and Jim A Schirillo. *Unifying multisensory signals across time and space.* Experimental Brain Research, vol. 158, no. 2, pages 252–258, September 2004.

[Wallace *et al.* 2006] Mark T Wallace, Brian N Carriere, Thomas J Perrault Jr, J William Vaughan and Barry E Stein. *The Development of Cortical Multisensory Integration.* Journal of Neuroscience, vol. 26, no. 46, pages 11844–11849, November 2006.

[Wallace 2004] Mark T Wallace. *The development of multisensory processes.* Cognitive Processing, vol. 5, no. 2, pages 69–83, June 2004.

[Watkins & Dayan 1992] Christopher J C H Watkins and Peter Dayan. *Q-learning.* Machine Learning, vol. 8, no. 3-4, pages 279–292, May 1992.

[Weiss *et al.* 2002] Yair Weiss, Eero P Simoncelli and Edward H Adelson. *Motion illusions as optimal percepts.* Nature Neuroscience, vol. 5, no. 6, pages 598–604, June 2002.

[Weisswange *et al.* 2009a] Thomas H Weisswange, Constantin Rothkopf, Tobias Rodemann and Jochen Triesch. *A Reinforcement learning model develops causal inference and cue integration abilities.* In Bernstein Conference on Computational Neuroscience (BCCN 2009), Frankfurt, 2009. Frontiers in Neuroscience.

[Weisswange *et al.* 2009b] Thomas H Weisswange, Constantin A Rothkopf, Tobias Rodemann and Jochen Triesch. *Can reinforcement learning explain the development of causal inference in multisensory integration?* In 8th International Conference on Development and Learning (ICDL 2009), pages 1–7, Shanghai, China, June 2009. IEEE.

[Weisswange *et al.* 2010] Thomas H Weisswange, Constantin A Rothkopf, Tobias Rodemann and Jochen Triesch. *Model averaging as a developmental outcome of reinforcement learning.* In Computational and Systems Neuroscience (COSYNE 2010). Frontiers in Systems Neuroscience, 2010.

[Weisswange *et al.* 2011] Thomas H Weisswange, Constantin A Rothkopf, Tobias Rodemann and Jochen Triesch. *Bayesian cue integration as a developmental outcome of reward mediated learning.* PLoS ONE, vol. 6, no. 7, page e21575, 2011.

[Wepsic 1966] James G Wepsic. *Multimodal sensory activation of cells in the magnocellular medial geniculate nucleus.* Experimental Neurology, vol. 15, no. 3, pages 299–318, July 1966.

[Werbos 1990] Paul J Werbos. *Backpropagation through time: what it does and how to do it.* Proceedings of the IEEE, vol. 78, no. 10, pages 1550–1560, 1990.

[Whiteley & Sahani 2008] Louise Whiteley and Maneesh Sahani. *Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes.* Journal of Vision, vol. 8, no. 3, pages 2.1–15, 2008.

[Wickens *et al.* 2003] Jeffery R Wickens, John N J Reynolds and Brian I Hyland. *Neural mechanisms of reward-related motor learning.* Current Opinion in Neurobiology, vol. 13, no. 6, pages 685–90, December 2003.

[Wickens 2009] Jeffery R Wickens. *Synaptic plasticity in the basal ganglia.* Behavioural Brain Research, vol. 199, no. 1, pages 119–28, April 2009.

[Wigström & Gustafsson 1986] H Wigström and B Gustafsson. *Postsynaptic control of hippocampal long-term potentiation.* Journal of Physiology-Paris, vol. 81, no. 4, pages 228–36, January 1986.

[Williams & Zipser 1989] Ronald J Williams and David Zipser. *A Learning Algorithm for Continually Running Fully Recurrent Neural Networks.* Neural Computation, vol. 1, no. 2, pages 270–280, June 1989.

[Wozny & Shams 2011] David R Wozny and Ladan Shams. *Recalibration of Auditory Space following Milliseconds of Cross-Modal Discrepancy.* Journal of Neuroscience, vol. 31, no. 12, pages 4607–4612, March 2011.

[Wozny *et al.* 2010] David R Wozny, Ulrik R Beierholm and Ladan Shams. *Probability Matching as a Computational Strategy Used in Perception.* PLoS Computational Biology, vol. 6, no. 8, August 2010.

[Wunderlich *et al.* 2011] Klaus Wunderlich, Ulrik R Beierholm, Peter Bossaerts and John P O'Doherty. *The human prefrontal cortex mediates integration of potential causes behind observed outcomes.* Journal of Neurophysiology, vol. 106, no. 3, pages 1558–1569, September 2011.

[Wysoski *et al.* 2010] Simei Gomes Wysoski, Lubica Benuskova and Nikola Kasabov. *Evolving spiking neural networks for audiovisual information processing.* Neural Networks, vol. 23, no. 7, May 2010.

[Xie & Seung 2004] Xiaohui Xie and H Sebastian Seung. *Learning in neural networks by reinforcement of irregular spiking.* Physical Review E, vol. 69, no. 4 Pt 1, page 041909, April 2004.

[Xu *et al.* 2005] Dongming Xu, Jing Lan and Jose C Principe. *Direct adaptive control: an echo state network and genetic algorithm approach.* In International Joint Conference on Neural Networks (IJCNN 2005), volume 3, pages 1483–1486. IEEE, 2005.

[Yamada *et al.* 2011] Hiroshi Yamada, Hitoshi Inokawa, Naoyuki Matsumoto, Yasumasa Ueda and Minoru Kimura. *Neuronal basis for evaluating selected action in the primate striatum.* European Journal of Neuroscience, vol. 34, no. 3, pages 489–506, July 2011.

[Yang & Shadlen 2007] Tianming Yang and Michael N Shadlen. *Probabilistic reasoning by neurons.* Nature, vol. 447, no. 7148, pages 1075–1080, June 2007.

[Yin *et al.* 2004] Henry H Yin, Barbara J Knowlton and Bernard W Balleine. *Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning.* European Journal of Neuroscience, vol. 19, no. 1, pages 181–189, January 2004.

[Young *et al.* 1993] Mark J Young, Michael S Landy and Laurence T Maloney. *A perturbation analysis of depth perception from combinations of texture and motion cues.* Vision Research, vol. 33, no. 18, pages 2685–2696, December 1993.

[Yu *et al.* 2010] Liping Yu, Benjamin A Rowland and Barry E Stein. *Initiating the Development of Multisensory Integration by Manipulating Sensory Experience.* Journal of Neuroscience, vol. 30, no. 14, pages 4904–4913, April 2010.

[Zaidel *et al.* 2011] Adam Zaidel, Amanda H Turner and Dora E Angelaki. *Multisensory Calibration Is Independent of Cue Reliability.* Journal of Neuroscience, vol. 31, no. 39, pages 13949–13962, September 2011.

[Zemel *et al.* 1998] Richard S Zemel, Peter Dayan and Alexandre Pouget. *Probabilistic interpretation of population codes.* Neural Computation, vol. 10, no. 2, pages 403–430, February 1998.

[Zhang *et al.* 1998] Li I Zhang, Huizhong W Tao, Christine E Holt, William A Harris and Mu-ming Poo. *A critical window for cooperation and competition among developing retinotectal synapses.* Nature, vol. 395, no. 6697, pages 37–44, September 1998.

Glossary

causal inference Including multiple possible underlying structures, that could generate the observations, into an inference process. 9–12, 16

drift diffusion model A computational model of (binary) decision making. Stochastic evidence for one or the other option at each timestep is integrated, and a decision is made if this variable reaches a threshold. The behaviour of this variable shows a biased random walk behaviour.. 6, 62

Hebbian learning The first theory of neuronal learning introduced in [Hebb 1949]. Its most basic form states "*Neurons that fire together, wire together*". In computational studies that usually means that a connection is potentiated if the input neuron is active and at the same (or close in) time also the output neuron fires.. 15, 16

Q-value The predicted reward value when in a certain state and performing a certain action. 22, 23, 25–27, 34

receptive field The response profile of a neuron along one or multiple dimensions of inputs. 12, 62

temporal difference error The difference between the reward prediction for a given state–action pair and the true reward received after performing that action in that state including possible future rewards from that state. 18, 62

Acronyms

2-AFC two-alternative forced choice. 7, 8, 25

AES anterior ectosylvian sulcus. 12, 15

AI model always integrating the cues. 24, 26, 29

AMPAR AMPA receptors. 15

ANN artificial neural network. 15, 18, 22, 25, 37

DDM drift diffusion model. 6, 12, 15

fMRI functional magnetic resonance imaging. 20

LTD long term depression. 20

LTP long term potentiation. 20

MA model averaging. 5, 10, 24, 26, 29, 30, 37

MAP estimate maximum *a posteriori* estimate. 5, 6

MS model selection. 5, 10, 24, 26, 29, 30

NI model always treating the cues as independent – never integrating. 24, 26, 29, 30

NMDAR NMDA receptors. 15

PM probability matching. 10, 24, 25, 30, 37

PPC probabilistic population code. 14

PSE point of subjective equality. 7, 8, 25, 26

RF receptive field. 12–16

RL reinforcement learning. 16, 18–20, 22, 24–26, 29, 30, 37, 38

SC superior colliculus. 11–13, 15

TD temporal difference. 18, 23, 37

VTA ventral tegmental area. 19, 20

Acknowledgments

I would like to thank my principle supervisor Jochen Triesch for the opportunity to do my PhD at FIAS, for helping defining my projects and supporting the implementation and interpretation of the results. I am also happy that he is hosting such a great diversity of projects in computational neuroscience and cognition, the main reason why I think I acquired a good general knowledge base in the field. I thank Professor Ramesh for accepting the extra work of being my second supervisor. I thank my co-advisor Constantin Rothkopf for his steady supervision, his impressive ability to produce names of researchers or papers relevant for the current problem, and for not taking all my jokes too seriously as well as for his support on navigating in the scientific world. I thank my co-advisor Tobias Rodemann for his supervision, especially during my time at the Honda Research Institute Europe. I thank him and Cem Karaoguz for sharing their data with me and for good collaboration on the adaption of the RL algorithm on this data.

I thank Cristina Savin for introducing me to the basic concepts of reservoir computing with plasticity and for providing the code of her own network from [Savin 2010]. In addition to my supervisors I also want to thank her as well as Daniel Krieg, Philip Sterne, Quan Wang and Pengsheng Zheng for discussions about various problems with the recurrent network. I thank my three supervisors as well as Daniel Krieg, Quan Wang and Daniela Pamplona for proofreading of the manuscript.

I want to thank Cristina Savin and my office mates Daniela Pamplona, Stephanie Lo, Jörg Bornschein and Christian Keck for their great company with many interesting scientific and non-scientific discussions and for many great memories. Generally I enjoyed most of my time as a PhD student due to all the great people in the Triesch Group and in the rest of FIAS neuroscience. I want to thank all of them for help and discussions and for the distractions from mensa food during lunch.

Thanks goes also out to my family and my friends for their great support at all times and for giving me many good reasons to get all the work done efficiently so I can spend time with them.

Finally I am thankful for the financial support by the German Federal Ministry of Education and Research (BMBF) within the "Bernstein Focus: Neurotechnology" through research grant 01GQ0840 and by the Honda Research Institute Europe GmbH and also for the administrative support of the FIAS staff, in particular Gaby Schmitz.