

## مدار مغز برای تصمیم‌گیری در راسته پستانداران غیر انسان

### مقدمه

مطالعات تجربی در مورد تصمیم‌گیری (در اغلب موارد) به بررسی انتخابها با احتمالات و نتایجی که به وضوح تعریف شده‌اند می‌پردازد که در آن تصمیم‌گیرندگان گزینه‌هایی را انتخاب می‌کنند که برای آنها عواقبی را در بر دارد. در واقع این امر نشان می‌دهد که وظایف تصمیم‌گیری متعارف شامل انتخاب بین شرط‌بندیهای پولی است، برای مثال ممکن است از شرکت‌کنندگان این پرسش بیان شود که آیا 50 درصد شانس را برای 25 دلار ترجیح می‌دهند یا شانس قطعی برای 10 دلار را ترجیح می‌دهند.

اگرچه نتایج و احتمالات غالباً پیچیده و نامشخص است و در بعضی مواقع با ابهام همراه است، این تصمیمات معمولاً در تنظیمات جداشده اجتماعی قرار می‌گیرند. با این حال، در تصمیمات روزمره ما به ندرت در این شرایط قرار می‌گیریم، و در حقیقت بسیاری از تصمیمات و انتخابها در بستر تعامل اجتماعی گزینش می‌شوند. ما زندگی می‌کنیم، کار می‌کنیم و در محیط بسیار پیچیده اجتماعی بازی می‌کنیم، و تصمیماتی که می‌گیریم غالباً به تصمیمات دیگران بستگی دارد، برای مثال، زمانی که ما در حال تصمیم‌گیری برای گسترش یک پیشنهاد شغلی هستیم یا زمانی که ما وارد یک مذاکره تجاری می‌شویم. این تصمیمات این پتانسیل را دارند که پنجره مفیدی را به داخل اشکال پیچیده تصمیم‌گیری باز کنند، تصمیماتی که تقریباً بسیاری از گزینه‌های جالب را در زندگی واقعی ما انتخاب می‌کند. این بخش به بررسی رویکرد اقتصاد عصبی برای مطالعه این مسئله در پستانداران غیر انسانی می‌پردازد، که با اندازه‌گیری مستقیم یا دستکاری سیگنالهای عصبی در میمون‌هایی که در چنین تصمیمات اجتماعی درگیر هستند صورت می‌گیرد.

ماهیت تصمیم‌گیری ممکن است اساساً هنگامی که نتیجه تصمیم‌گیری به تصمیمات دیگران بستگی دارد تغییر کند، موضوعی که در فصل‌های 2، 11، 25 و 27 این کتاب نیز آورده شده است. تحت این نوع از شرایط، محاسبه ابزار مورد انتظار استاندارد که بسیاری از نظریه‌های موجود را پایه‌ریزی می‌کند و مدل‌های تصمیم‌گیری توصیف شده در فصل 1 با این حقیقت که ما باید برای فهمیدن اعتقادات شریک یا رقیب خودمان که در تلاش برای رسیدن به تصمیم بهینه است تلاش کنیم را پیچیده کرده است، همانطوری که در فصل 2 نیز به آن اشاره شده است.

بعنوان بخشی از رویکرد اقتصاد عصبی، چندین گروه از محققان شروع به بررسی روابط روانی و عصبی تصمیمات اجتماعی نسبتاً ساده با استفاده از وظایف مشتق شده از شاخه‌ای از اقتصاد تجربی که بر روی تئوری بازی تمرکز دارد کرده‌اند. این وظایف و امور، در عین سادگی ممکن است نیازمند استدلال پیچیده‌ای در مورد انگیزه‌های دیگر بازیکنان در این کار باشد. ترکیب این امور و روش‌های اعصاب‌شناختی تا حد زیادی توانایی بالایی برای بسط دادن دانش مکانیسم‌های ذهنی موجود در تصمیم‌گیری اجتماعی و نیز پیشرفت مدل‌های نظری تصمیم‌گیری در یک محیط اجتماعی غنی را دارد.

این فصل بر روی استفاده از روش‌های الکتروفیزیولوژیک تهاجمی در میمون‌ها برای مطالعه فرآیند تصمیم‌گیری در طول فعالیت‌های نظری-بازی تمرکز دارد. بزرگترین مزیت این روش این است که به ما اجازه می‌دهد که بطور مستقیم سیگنال‌ها و مدارهای عصبی را با تفکیک‌پذیری فضایی و زمانی مطلوب در طول فرآیند تصمیم‌گیری اندازه‌گیری و دستکاری کنیم. به دلیل شماری از دلایل فنی که در زیر مورد بحث قرار خواهد گرفت، این رویکرد به بازیهای تکراری ساده مانند سنگ-کاغذ-قیچی محدود شده است. این بازیهای ساده برای بررسی فرآیندهای عصبی موجود در ارائه پاداش، احتمال، ارزش ذهنی، انتخاب گزینه و یادگیری سازگار ایده‌آل هستند. خواننده به فصل‌های 11 و 25 که رویکرد اقتصاد عصبی عادی تصویربرداری مغز در انسانها در طول بازیها ارائه شده است ارجاع داده می‌شود. مزیت تصویربرداری مغز انسان این است که به بررسی فرآیندهای تصمیم‌گیری در گونه‌هایی که بیشتر مورد علاقه ماست می‌پردازد. همچنین، بیشتر بازیهای پیچیده‌تر می‌توانند برای بررسی ترجیحات اجتماعی و مفاهیم وابسته در انسانها به کار برده شوند مانند عدالت، متقابل به مثل و اعتماد که نقش مهمی را در به چالش کشیدن موقعیت‌های اجتماعی ایفا می‌کنند.

تکنیکهای تصویربرداری عصبی که در حال حاضر موجود هستند فاقد غنای فضایی و زمانی اندازه‌گیریهای نوروفیزیولوژیک مستقیم هستند. در مجموع، این رویکردهای انسانی و غیر انسانی برای ما دسترسی به فرآیند موجود در جعبه سیاه و درک عمیقتر اینکه چگونه حیوانات اجتماعی به موفقیت (و گاهی با عدم موفقیت) با هم تعامل می‌کنند را فراهم می‌آورد.

## تئوری بازی

در اصل، تئوری بازی مجموعه‌ای از مدل‌های سختگیرانه با هدف توضیح موقعیتهایی است که در آن تصمیم‌گیرندگان باید با یکدیگر تعامل داشته باشند و این موضوع در فصل 2 این کتاب به تفصیل آمده است. در تئوری بازی کلاسیک (برای مثال، ون نیومن و مورگن اشترن، 1944) فرض بر این است که تصمیم‌گیرندگان دانش کاملی نه تنها در مورد هر کدام از اقدامات متناوبی که می‌توانند انتخاب کنند دارند، بلکه همچنین در مورد اینکه چگونه بازپرداخت بطور مشترک با اقدامات آنها و دیگر تصمیم‌گیرندگان تعیین می‌شوند نیز مطلع هستند. مفهوم تعادل نقش مهمی را در فهم این تعاملات دارد. برای مثال، مجموعه‌ای از استراتژی‌ها بعنوان تعادل ناش یعنی زمانی که هیچ بازیکن فردی نمی‌تواند بازپرداخت را با انحراف از چنین استراتژی‌هایی بصورت یکطرفه افزایش دهد ارجاع داده می‌شوند (ناش، 1950، همچنین فصل 2 را نیز ببینید). بعنوان مثال، اگر دو بازیکن در یک بازی سنگ-کاغذ-قیچی در حال انتخاب از بین سه گزینه غیرقابل پیش‌بینی و در نسبت‌های مساوی باشند (یک استراتژی مخلوط)، ممکن است در تعادل ناش باشند زیرا هیچکدام از آنها انگیزه‌ای برای تغییر استراتژی خود ندارند. چنین تعادل نظری بازی باعث دقیق شدن مدل‌های تصمیم‌گیری انسانی و غیر انسانی خواهد شد، با این حال، تنها تا حدی که تصمیم‌گیرندگان واقعی قادر به ساختن چنین استنتاج‌هایی هستند، شناسایی و پیاده‌سازی چنین استراتژی‌هایی ضروری به نظر می‌رسد. در حقیقت، زمانیکه رفتارهای انسان و حیوان در طول بازی‌های مختلف به صورت سیستماتیک در آزمایشگاه مورد مطالعه قرار می‌گیرد، نتایج غالباً انحرافات سیستمی مشابهی را از پیش‌بینی‌های استراتژی‌های تعادل نشان می‌دهد (نکته‌ای که در فصل قبل و در مطالعه کارمر در سال 2013 مطرح شده است). بطور معمول، تصمیم‌گیرندگان خودخواهی کمتری دارند و بیشتر تمایل به در نظر

گرفتن عواملی مانند تقارن و عدالت به جای نظریه بازی کلاسیک دارند (فصل 11). علاوه بر این، زمانیکه همان بازی بصورت مکرر تکرار شود، تصمیم‌گیرندگان تمایل دارند که استراتژیهای خود را به تدریج برای بهبود نتایج انتخاب خود تطبیق دهند. در واقع، انسانها و میمونها پویایی مشابهی را در رفتارهای انتخابی خود در طول بازیهای تکراری نشان می‌دهند (لی، 2008)، و این به نحوی است که غالباً در نظریه بازی کلاسیک گنجانده نشده است. با این حال برای خواننده مهم است که در نظر داشته باشد که به رغم این شباهتهای استراتژیک بین گونه‌ها، معلوم نیست که آیا میمونهایی که آزمایشات را در آزمایشگاهها انجام می‌دهند به واقع این را می‌فهمند که آنها در بازی استراتژیک درگیر هستند یا محققان می‌توانند برای آنها آموزش کلامی یا دریافت گزارش از آنها را فراهم کنند، زیرا آنها غالباً با یک رقیب زنده در آزمایشگاه مواجه نیستند.

با این وجود، وظایف مشخص و رویکرد مدلسازی رسمی ارائه شده توسط تئوری بازی یک بنیاد مفید را برای مطالعه تصمیمات در بستر اجتماعی فراهم می‌سازد. از نقطه نظر تجربی، چارچوب ریاضی تئوری بازی یک زبان مشترک را فراهم می‌آورد که در آن یافته‌های گروههای تحقیقاتی مختلف و در واقع روشهای تحقیقاتی می‌توانند مقایسه شوند و انحرافات از پیش‌بینی‌های مدل نیز تعیین گردند. این وظایف یک الگوی غنی و متغیر از تصمیم‌گیری را با استفاده از قوانین بسیار ساده ارائه می‌دهند. قوانین برای سه بازی تکراری که در میمونها (تطبیق پول، بازی بازرسی و سنگ، کاغذ، قیچی) مطالعه شده است به شکل معمول در شکل 26.1 نشان داده شده است. همانطور که نتایج را به شرح زیر توصیف می‌کنیم، روندی را مورد بررسی قرار می‌دهیم که در آن این فعالیتها و وظایف برای آزمایشات نوروفیزیولوژیک در بیداری و رفتار میمونها تطبیق داده شده است.

(A) **Matching pennies game**

		'Evens' - Monkey	
		Heads	Tails
'Odds' - Computer	Heads	1	0
	Tails	0	1

(B) **Inspection game**

		'Inspector' - Computer	
		Inspect	Not inspect
'Worker' - Monkey	Certain	2-1	2
	Risky	1-1	0

(C) **Rock-paper-scissors game (Traditional)**

		Subject #1		
		Rock	Paper	Scissors
Subject #2	Scissors	1	-1	0
	Paper	-1	1	0
	Rock	0	1	-1

(D) **Rock-paper-scissors game (Biased)**

		Computer		
		Rock	Paper	Scissors
Monkey	Scissors	0	-4	-1
	Paper	-3	-1	0
	Rock	3	1	0

شکل 26.1. ماتریس‌های بازپرداخت برای بازیهای با استراتژی مخلوط مورد استفاده در آزمایشات نوروفیزیولوژیک در راسته غیر انسان. فعالیت تطبیق پول (آ)، فعالیت بازی بازرسی (ب)، فعالیت سنگ-کاغذ-قیچی سنتی (ج) و فعالیت سنگ-کاغذ-قیچی مغرضانه (د). اعداد قرمز در هر سلول به واحدهای پاداش دریافت شده توسط میمون اختصاص دارد (یا واحدهای مجازی برای اعداد آبی در مورد رقیب کامپیوتری). برای فعالیت بازی بازرسی (ب)،  $a$  به هزینه بازرسی برای رقیب کامپیوتری بر می‌گردد و در محدوده 0.1 تا 0.9 در امتداد بلوکهای آزمایش در 0.2 افزایش یافته‌اند.

مزایای ترکیب کردن فعالیتهای نظریه بازی با تکنیک‌های علوم اعصاب مانند ثبت تک نرونی دارای دو قسمت است. اول، همانطور که در بالا بحث شد، الگوهای انتخاب در این فعالیتهای غالباً مطابق با پیش‌بینی‌های نظریه بازی کلاسیک نیست، و بنابراین توصیف دقیق‌تر رفتار (از لحاظ فرآیند عصبی که پایه آنها است) در تطبیق این مدلها برای متناسب

کردن اینکه چگونه تصمیمات گرفته می‌شوند حائز اهمیت است. دوم، علم اعصاب می‌تواند محدودیت‌های بیولوژیکی مهم را در فرآیندهای درگیر فراهم سازد، و در واقع تحقیقات نشان می‌دهد که بسیاری از فرآیندهای که فکر می‌کنند پایه این نوع از تصمیم‌گیری پیچیده باشند ممکن است قویا با بیشتر فرایندهای مغزی مانند پاداش، تنبیه و انزجار همپوشانی داشته باشند. دانش "بلوکهای ساختمانی" تصمیم‌گیری در بازیها بطور قابل توجهی در ساخت مدل‌های بهتر این فرآیند کمک خواهد کرد.

## مدار بصری راسته پستانداران غیر انسان بعنوان یک سیستم مدل برای مطالعه پایه عصبی تصمیم‌گیری اجتماعی

با وجود استفاده از بیداری، رفتار میمون‌ها یکی از اصلی‌ترین سیستم‌های تحقیقاتی علوم اعصاب در چهل سال اخیر بوده است و استفاده از آنها در ارتباط با فعالیتهای تئوری بازی کمتر از 10 سال است. اگرچه هنوز در به بلوغ کامل نرسیده است، این تحقیق بینش قابل توجهی را در مورد فرآیندهای پنهانی که در جعبه سیاه در طول تعاملات اجتماعی رخ می‌دهند ایجاد کرده است. ما در اینجا روشهای نوروفیزیولوژیک را برای تکنیکهای غیر عصبی معرفی می‌کنیم و محدودیتهای آنها را در فراهم آوردن بینشهای آتی برجسته می‌سازیم.

یک مدل حیوان مناسب لازم است تا دسترسی مستقیم به بستر عصبی در طول تصمیم‌گیری در بازیها فراهم شود. به دلایل مختلف، میمون رزوس (ماکاکا مولاتا) مدل حیوانی اصلی برای مطالعه فرآیندهای تصمیم‌گیری مرتبه بالا بوده است. سازماندهی عمومی سیستم عصبی آنها شبیه انسان است و این پیچیدگی به آنها این اجازه را می‌دهد که فعالیتهای رفتاری نسبتا پیچیده را در آزمایشگاه یاد بگیرند. مناسب بودن این راسته از پستانداران احتمالا به حوزه اجتماعی نیز می‌رسد، نکته‌ای که در فصل 7 آمده است. هر دو گونه بخوبی دارای ساختارهای اجتماعی با سیستم‌های سیگنال‌دهی پیچیده برای حفظ این ساختار هستند (بیرن و وایتن، 1989؛ د وال، 1990). در شماری از بسترهای تصمیم‌گیری شامل بازیهای با استراتژی مخلوط که ما در اینجا بر روی آنها تمرکز داریم، میمون‌ها و انسانها استراتژیهای

قابل مقایسه‌ای را نشان می‌دهند که نشان دهنده این است که ممکن است بسیاری از فرآیندهای عصبی به اشتراک گذاشته شوند.

به علت تعدادی از دلایل عملی، تحقیق در مورد تصمیم‌گیری در حیوانات مورد توجه قرار گرفته است، اما نه بصورت انحصاری بر روی سیستم بصری-عصبی میمون (گلیمچر، 2003؛ شال و تامسون، 1999). سیستم بصری-عصبی پستانداران غیر از انسان حائز اهمیت است زیرا به آنها اجازه می‌دهد که به اندازه کافی اطلاعات بصری را از محیط ما استخراج کنند. این امر با همتراز کردن فرورفتگی (بخش مرکزی سبکیه که با بالاترین حد بینایی همراه است) با اهداف مورد نظر با استفاده از حرکات چشم که بعنوان ساکاد شناخته می‌شود (و سپس ثابت می‌شود) زمانی که اطلاعات بصری کسب می‌شود و در نواحی مخطط مورد پردازش قرار می‌گیرد حاصل می‌شود. اگرچه انتخابها بصورت سنتی در نظر گرفته نمی‌شوند، ساکادها در حقیقت عبارتند از بازخوانی رفتاری یکی از رایجترین تصمیمات ما یعنی انتخاب اینکه چه زمانی و چه جایی را نگاه کنیم.

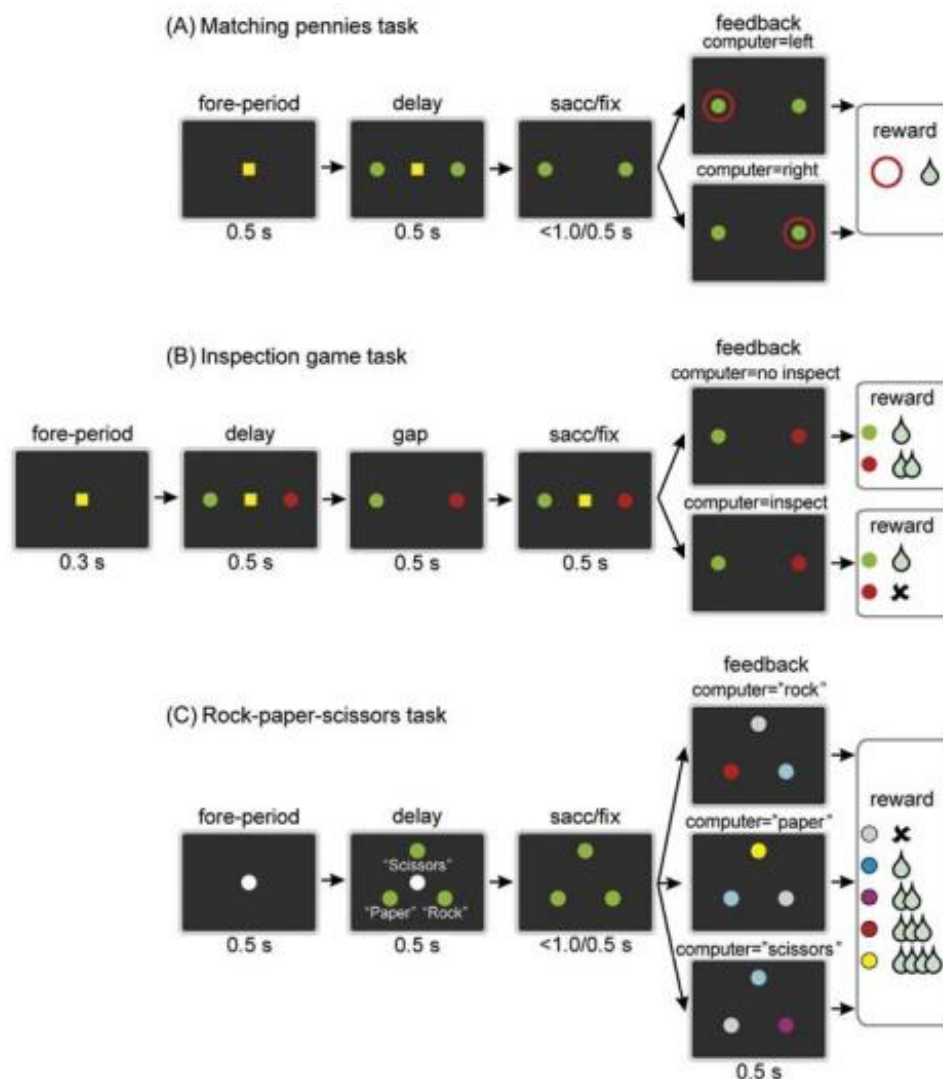
مدار عصبی اساس فرآیند بصری است و کنترل حرکت چشم بخوبی درک می‌شود، و یک بنیاد جامع را برای پرسیدن سوالات در مورد فرآیندهای تصمیم‌گیری که حساسیت را به عمل پیوند می‌دهند فراهم می‌سازد. حرکات چشمی، ساده و دارای حرکات کلیشه‌ای هستند و برخلاف دیگر سیستم‌های حسی-حرکتی، تمام مدار عصبی آن درون سر قرار گرفته است. این نکته آخر حائز اهمیت است زیرا سر می‌تواند در طول آزمایشات از حرکت باز بایستد و بنابراین پایداری مورد نیاز را برای ثبت نرون‌های کوچک در حیوان در حال حرکت یا خواب را فراهم می‌کند. برای انجام این کار، میمونها برای نشستن در صندلی‌هایی که سر را در طول انجام آزمایشات از حرکت محدود می‌کنند آموزش می‌بینند. در نتیجه، تعاملات اجتماعی در آزمایشگاه شامل هدایت حرکات چشم به سمت اهداف بصری است که توسط رقبای کامپیوتری مجازی به جای تعاملات مستقیم، خشن و آشفته بین میمونها کنترل می‌شود (شکل 26.2). محدودیتهای قابل مقایسه هنگامی که آزمایشات بر روی تصمیم‌گیری اجتماعی در اسکنرها در آزمایشات انسانی انجام می‌شوند وارد می‌گردند.

## مزایا و معایب رویکرد نوروفیزیولوژی سیستم‌ها

مزایای رویکرد نوروفیزیولوژی از دسترسی مستقیم به بستر عصبی ناشی می‌شود. سیگنال‌های عصبی می‌توانند با تفکیک‌پذیری زمانی (1 میلی ثانیه) و فضایی (نورونهای منفرد) دقیق و با دقت قابل مقایسه نمونه‌برداری شوند، فعالیت عصبی می‌تواند به صورت مصنوعی دستکاری شود.

برای آنهایی که با روش همراه با نوروفیزیولوژی در بیداری و رفتار میمون‌ها آشنایی ندارند، ما بطور خلاصه شرح می‌دهیم. این مورد بطور دقیقتر و با جزییات در فصل 6 بیان می‌شود. یک محفظه با کلاهک قابل جابجایی در یک ورودی کوچک در جمجمه تحت شرایط ضد عفونی قرار داده شده است. در شروع هر آزمایش، یک الکتروود فلزی یا سوزن وارد غشا می‌شود که مغز را پوشش می‌دهد و به اهستگی تا ناحیه مغزی مورد نظر پایین آورده می‌شود. این روش‌ها بدون درد هستند و آسیب کمی را به بافت عصبی وارد می‌کنند زیرا مغز فاقد گیرنده‌های درد است و تنها کاوشگرهای نازک استفاده می‌شوند. این خصوصیات بحرانی هستند زیرا برای کسب نتایج آزمایشگاهی دقیق حیوان و مغز باید تا حد امکان در حالت طبیعی خود باشند.





شکل 26.2. فعالیت تطبیق پول (آ) فعالیت بازی بازرسی (ب) و فعالیت سنگ-کاغذ-قیچی (ج) مورد استفاده برای مطالعه رفتارها در میمونها. (آ) در طول فعالیت تطبیق پول، حیوانات گزینه‌های خود را با نگاه خود به سمت یکی از دو هدف جانبی نشان دادند. در طول دوره بازخورد، هدف انتخاب شده با رقیب کامپیوتری توسط یک حلقه قرمز نشان داده شد و حیوان نیز تنها زمانی که همان هدف را انتخاب می‌کرد پاداش داده می‌شد. (ب) در طول فعالیت بازی بازرسی، حیوانات گزینه‌های خود را بعد از یک گپ زمانی با تغییر نگاه خود به سمت هدف قرمز خطرناک در محدوده پاسخ نورون یا هدف سبز در مقابل محدوده پاسخ نورون نشان دادند. میمون همواره یک واحد آب را برای انتخاب هدف خاص دریافت می‌کرد. زمانی که میمون هدف خطرناک را انتخاب می‌کرد، 2 واحد آب را در صورت بازرسی نکردن کامپیوتر دریافت می‌کرد و اگر رقیب کامپیوتری بازرسی می‌کرد صفر واحد آب دریافت می‌کرد. (ج) در

طول فعالیت سنگ-کاغذ-قیچی، لازم بود که حیوانات نگاه خود را به سمت یکی از سه هدف محیطی اطراف تغییر دهند. در طول دوره بازخورد، نتیجه واقعی از هدف انتخاب شده و نتایج فکری از دیگر اهداف انتخاب شده که با بازی سنگ-کاغذ-قیچی مغرضانه تعیین شده بودند توسط رنگهای مختلف مشخص شدند.

این پتانسیل‌های عمل یا پالسهای الکتریکی ناشی از یک نورون و تکثیر در طول آکسون‌های توصیف شده در فصل 5 هستند که با استفاده از میکروالکترودها در طول این آزمایشات ثبت شدند. این فعالیت عصبی می‌تواند با خصوصیات دستورالعمل حسی، متغیرهای داخلی پیش‌بینی شده توسط نظریه اقتصادی، جوانب پاسخ انتخاب و نوع تقویت همبستگی داشته باشند. از آنجا که این فعالسازی عصبی می‌تواند با دقت میلی‌ثانیه اندازه‌گیری شود، بهترین ابزار برای فهم فهم محاسبات لحظه به لحظه است که اساس فرآیند تصمیم‌گیری می‌باشد.

دستکاری مصنوعی فعالیت عصبی می‌تواند شواهد مستقیم بیشتری را فراهم کند که نشان دهد ناحیه مغزی در فرآیند تصمیم‌گیری دخیل است. تعدادی از تکنیکها برای دستکاری فعالیت عصبی در فصل 6 توصیف شده‌اند. این فصل به توصیف تهییج مصنوعی فعالیت عصبی از طریق میکرو تحریک الکتریکی می‌پردازد. دقت زمانی، وسعت فضایی و شدت دستکاری فعالیت عصبی می‌تواند با دقت بیشتری با استفاده از تکنیکهای موجود برای غیرفعال کردن غیربرگشتی نواحی مغز کنترل شود.

تعدادی از معایب بالقوه موجود در استفاده از راسته پستانداران غیر انسانی برای تشخیص فرآیندهای عصبی پایه تعاملات اجتماعی انسان هستند. تا به امروز، راسته پستانداران غیر انسانی برای اجزای بازبهای با استراتژی مخلوط ساده در طول ثبت نوروفیزیولوژیکی آموزش دیده‌اند. خواننده باید به فصلهای 2، 7، 11 و 25 برای بحث در مورد دیگر اشکال تعاملات اجتماعی رجوع کند. بسیاری از چالش موجود در استفاده از راسته پستانداران غیر انسان این امر را ارزیابی می‌کند که آیا آنها توانایی‌های شناختی کلیدی را با ما برای اجرای تعاملات اجتماعی پیچیده به اشتراک می‌گذارند، و اگر چنین است تقطیر این وظایف انتزاعی به شکلهایی که می‌توانند درک کنند صورت می‌گیرد. علاوه بر این، ممکن است آموزش حیوانات بر روی فعالیتهای تئوری بازی بدون دستورالعمل کلامی و تنها با استفاده از تکنیکهای عملیاتی مشکل باشد.

حتی اگر استراتژیهای انتخاب قابل مقایسه در طول بازیهای آزمایشی استفاده شوند، ما باید به خاطر داشته باشیم که این پیش‌نیاز است نه اثبات که همان مکانیسم‌های عصبی در این دو گونه به اشتراک گذاشته می‌شوند. گفته می‌شود که میمون‌ها و انسانها استراتژیهای مشابه قابل توجهی را تحت بازیهای استراتژی مخلوط نشان داده‌اند. اگرچه محدودیت‌های این مدل حیوانی هنوز مشخص نشده است، فهم مکانیسم‌های عصبی که پایه تصمیم‌گیری در طول بازیها در میمون‌ها هستند مهم است زیرا اینها ممکن است بطور مستقیم با مکانیسم‌های تصمیم‌گیری ما در ارتباط باشند یا حداقل مکانیسم اصلی را که از طریق آن باقی فرآیندهای تصمیم‌گیری پیچیده‌تر را ارائه دهد.

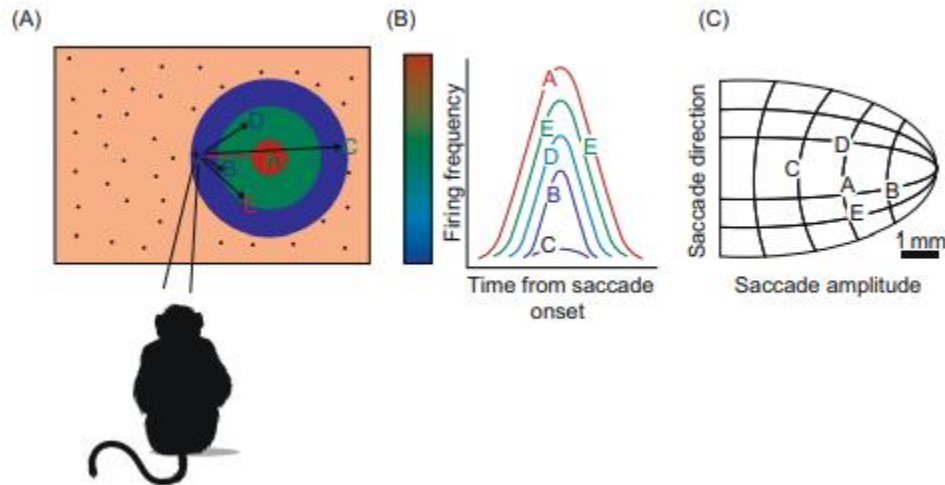
### انطباق بازیها برای راسته پستانداران غیر انسانی

عصب‌شناسان در ابتدا تلاش خود را بر روی بازیهای با استراتژی مخلوط ساده متمرکز کرده‌اند زیرا راسته پستانداران غیر انسان می‌تواند بطور نسبی و به آسانی بر روی این فعالیتها آموزش ببینند. سپس ما بطور مختصر برخی از این بازیها را توصیف می‌کنیم و اینکه چگونه آنها برای آزمایشگاه عصب اصلاح شده‌اند (شکل 26.2).

تقریباً تمامی فعالیتهای مطالعه شده تا به امروز شامل حیوانات تشنه و مقایسه آنها با رقبای کامپیوتری پویا برای پاداش آب بوده است. میمون‌هایی که در مقابل نمایش بصری مینشینند، انتخابهای خود را با نگاه کردن به یکی از چند هدف انتخابی تعیین می‌کنند که به دنبال آن بازخورد بصری انتخاب رقیب کامپیوتری است. در شروع هر آزمایش، یک میکروالکتروستیمولاسیون می‌شود که به عقب و جلو حرکت می‌کند و این امر تا زمانیکه آزمایش‌کننده موفق به جدا کردن فعالیت یک نورون منفرد از پس زمینه فعالیت مغز شود ادامه پیدا می‌کند.

یک مفهوم بحرانی برای تفسیر یافته‌ها مربوط به ناحیه پاسخ عصبی است. در بسیاری از نواحی مغزی درگیر در دید و کنترل حرکت چشم، هر نورون با ترکیب خاصی از حس و خصوصیات حرکتی که با هم ناحیه پاسخ نورون را تعریف می‌کنند فعال می‌شود. در برخی از ساختارها، مانند قشر بصری یا برجستگی زبرین، جمعیت نورونها با نواحی پاسخ مشابه بصورت یکپارچه به نقشه‌های توپوگرافی حس و فضای حرکت سازمان‌دهی می‌شوند (شکل 26.3). ویژگیهای حسی شامل مکان فضایی محرک بصری نسبت به فرورفتگی، سرعت و جهت حرکت، رنگ و شکل است. خصوصیات

حرکتی شامل جهت و دامنه حرکت چشم و زمان پاسخ است. اگر نورون‌ها در یک ناحیه مغزی دارای نواحی پاسخ با خصوصیات حسی و حرکتی تعریف شده باشند، آزمایشگر خواص ناحیه پاسخ نورون را تعیین می‌کند و اهداف انتخابی را برای مشغول کردن نورون تحت مطالعه در نظر می‌گیرد.



شکل 26.3. نمونه ناحیه پاسخ از یک نورون منفرد و نقشه‌های توپوگرافی جمعیتی در نواحی مغز با پاسخهای بصری و حرکتی. پاسخهای حرکتی چشم در فرورفتگی زبرین در اینجا توضیح داده شده است اما اساس کار شبیه پاسخهای بصری در بیشتر نواحی بصری است. (آ) هنگامی که فعالیت از نورون منفرد از نويز پس زمينه جدا می‌شود، میمون چشمش را از محل مرکزی به سمت اهداف موجود در ناحیه بصری (نقاط سیاه) تکان می‌دهد. (ب) مقدار فعالیت عصبی استخراج شده از این نورون برای پنج جهت مختلف در پنل آ نشان داده شده است. از این فعالیت‌های برداری یک نقشه حرارتی ایجاد می‌شود که نشان دهنده ناحیه پاسخ عصبی است. بردار حرکتی که با هدف آ همراه است بعنوان مرکز این ناحیه پاسخ نورون در نظر گرفته می‌شود زیرا بالاترین فرکانس را ایجاد می‌کند. (ج) نورون‌های داخل لایه‌های میانی بخش حرکتی بعنوان نقشه توپوگرافی بردارهای حرکتی سازمان داده می‌شوند. جمعیت نورون-ها با بالاترین فعالیت در سمت چپ نشان داده شده است که با پنج حرکت به سمت راست در پنل آ همراه است. نواحی پاسخ در قسمت‌های مختلف مغز بیشتر به دو روش که مربوط به فرآیند تصمیم‌گیری هستند توضیح داده می‌شوند. اول، خواص ناحیه پاسخ زمانی که ما از حس به حرکت مربوط به نواحی مغزی حرکت می‌کنیم نمود می‌کند، در نواحی قشری مغز، نواحی پاسخ خواص محرک را صرف نظر از حرکات یا تصمیمات گرفته شده رمزگذاری می‌کند

و بعداً در نواحی حرکتی نواحی پاسخ خواص حرکات را صرف نظر از خصوصیات حسی حاصل رمزگذاری می‌کنند. این تحول به خوبی با تحقیقات علم عصب در دهه‌های اخیر شناسایی شده است. دوم، فعالسازی ناحیه پاسخ در بسیاری از نواحی مغز با عوامل اقتصادی و شناختی شکل می‌گیرد حتی زمانی که ویژگیهای حسی و حرکتی ثابت شوند. این فرآیندهای مدولاسیون از تعاملات با دیگر نواحی مغز که فاقد نواحی پاسخ کلاسیک مانند بسیاری از بخشهای جلویی و قشری غده عصبی اصلی است ناشی می‌شود. پاسخهای عصبی از این نواحی تمایل دارند که ناهمگن باشند و حتی بصورت ضعیفی برای شاخصه‌های حرکتی و حسی فعالیت تنظیم شده‌اند. در زیر ما رویکردهای اقتصاد عصبی بیان می‌کنیم تا تعیین کنیم که چگونه متغیرهای اقتصادی مانند انتخابها، پاداش، ارزش ذهنی و اعتقادات در نواحی بالاتر قشر سر تا نواحی پاسخ حسی-حرکتی کلاسیک برای انتخاب اقدامات اجتماعی مناسب ارائه می‌شوند.

در یک آزمایش اقتصاد عصبی در میمونها، هر بازی یا دور با حیوانی که در حال تثبیت یک محرک بصری اصلی است شروع می‌شود. حیوان انتخابش را با حرکت دادن نگاه خود به یکی از اهداف محیطی بعد از یک مکث کوتاه نشان می‌دهد. این که حیوان یک پاداش مایع (یا هر چیز دیگری) دریافت کند به انتخاب آنها و رقیب کامپیوتری بستگی دارد. اگرچه الگوریتم‌های کامپیوتری در جزییات با هم تفاوت دارند، در طول بازیهای با استراتژی مخلوط رقابتی همه به دنبال الگوهایی در تاریخ حیوان در انتخابها و پاداشها برای پیش‌بینی و مقابله با اقدامات پیش رو هستند، بطوریکه آنها می‌توانند رقیب طبیعی یا بیولوژیکی را تقریب بزنند.

میمونها برای اجرای بازیهای مجموع-صفر مانند تطبیق پول و سنگ-کاغذ-قیچی و بازیهای غیر مجموع-صفر مانند بازی بازرسی آموزش دیده‌اند. دیگر ابزارهای موفق برای مطالعه تصمیم‌گیری انطباقی در راسته پستانداران غیر انسان از فعالیتهای کاوشی استفاده می‌کند که نتایجی را مطابق با قانون تطابق هراشتین در اختیار قرار می‌دهد. در طول این فعالیتهای تواتری که با آن حیوان یک هدف را انتخاب می‌کند به سمتی تمایل دارد که پاداش قطعی از آن هدف بدست می‌آید. از آنجا که تطبیق فعالیتهای این قانون شامل تعامل با رقیب استراتژیک نمی‌شود بنابراین آنها از لحاظ فنی بازی محسوب نمی‌شوند. با این اوصاف، ما آنها را اینجا ذکر می‌کنیم زیرا مشخص نیست که آیا میمونها می‌توانند بین

این دو نوع از فعالیتهای تطبیقی تمییزی قائل شوند یا خیر. در واقع، آزمایشات قانون تطبیق و بازیهای سنتی ممکن است که بیشتر از آن چیزی که تایید شده است شباهت داشته باشند.

### تجزیه و تحلیل آماری داده‌های بازی تکراری

مزیت نهایی مطالعه تصمیم‌گیری اجتماعی در رفتار بیداری میمونها این است که سرمایه و ارزش انتخاب و داده‌های عصبی می‌توانند از هر بخش از آزمایش جمع‌آوری شوند. بطور معمول، میمونها صدها (اگرچه بیش از هزار نمی‌شود) آزمایش تکراری را در طول یک بخش آزمایشی منفرد انجام می‌دهند. این امر به چند دلیل یک مزیت محسوب می‌شود. اول، سیگنالهای عصبی و توالی انتخاب بسیار تصادفی هستند، بنابراین مجموعه داده‌های عظیم به شدت برای توسعه ارائه دقیقتر مشارکت نورون در استراتژی انتخاب کلی ارزشمند هستند. دوم، داشتن توالی‌های طولانی هر دو سیگنالهای عصبی و الگوهای انتخاب به محقق این اجازه را می‌دهد که به بررسی این که چگونه تاریخچه انتخابهای قبلی و نتایج آنها بر روی پردازش آزمایش کنونی تاثیر می‌گذارد بپردازد. این به خصوص برای پیگیری چنین عواملی مانند انتخابهای خود شخص و نتایج آنها، انتخابهای رقیب شما و نرخ پاداش کلی در طول تصمیم‌گیری اجتماعی حائز اهمیت است. اینها هر دو برای فراهم‌سازی تخمینهای دقیق ارزش ذهنی گزینه‌ها برای هدایت انتخاب فعلی حیاتی هستند و همچنین برای فرآیند یادگیری و تطبیق رقبا و شرایط پویا نیز ضروری هستند. در نهایت، چنین مجموعه داده‌های عظیمی به ما این اجازه را می‌دهند که مقایسات سختگیرانه‌ای از مدل‌های آماری مختلف برای فعالیت عصبی و انتخاب انجام دهیم. ما می‌توانیم این پرسش را مطرح کنیم که آیا نورون‌ها در ناحیه مغزی خاصی متغیرهای خاص پیش‌بینی شده توسط مدل‌های اقتصادی را نشان می‌دهند یا به تعیین اینکه کدام مدل‌های رقابتی بهترین توصیف را از یادگیری، رفتار انتخاب و فعالیت عصبی می‌پردازند.

با توجه به مقدار عظیمی از داده‌های انتخاب که می‌توانند از بخشهای متعدد آزمایشات رفتاری در میمونها حاصل شوند، تعدادی از مطالعات به مقایسه مدل‌های یادگیری مختلف پرداخته‌اند تا بینشی را در مورد ماهیت یادگیری که در طول بازیهای تکراری رخ می‌دهد کسب کنند. همانطور که در بخشهای زیر خلاصه می‌شود، این مطالعات همچنین به

شناسایی سیگنالهای عصبی در نواحی مغزی متعدد شامل قشر پیشانی و غده عصبی اصلی که به احتمال زیاد نقش مهمی را برای تصمیم‌گیری در طول تعاملات اجتماعی ایفا می‌کنند می‌پردازند.

## آموزش تقویتی

### آموزش تقویتی در بازیها

زمانی که تصمیم‌گیرندگان این اجازه را پیدا می‌کنند که بارها در بازی خاص تصمیم‌گیری کنند و نتایج انتخابهای خود و نیز انتخابهای دیگر بازیکنان را ببینند، رفتارهای آنها می‌تواند با مدل‌های یادگیری مختلف با دقت بیشتری نسبت به پیش‌بینی‌های تعادلی تئوری بازی کلاسیک توصیف شود. مدلها در تئوری یادگیری تقویتی با موفقیت توضیحات دقیقی را برای محدوده وسیعی از رفتارهای انتخابی (فصل 15 و 16) شامل آنهایی که در طول تعاملات اجتماعی رخ می‌دهند در اختیار قرار می‌دهد. تئوری یادگیری تقویتی تعداد زیادی از الگوریتم‌های محاسباتی را فراهم می‌ساز که می‌تواند برای کشف استراتژیهای موفق توسط سعی و خطا استفاده شود. در مقابل استراتژی تعادلی ایستا که توسط رویکردهای اقتصادی سنتی توصیف می‌شود، این مدل‌های یادگیری پیش‌بینی‌هایی را در مورد پویایی رفتار انتخابی آزمون به آزمون تولید می‌کنند. هدف چنین الگوریتم‌هایی این است که مجموع پاداشهای آتی را که معمولا با توجه به تاخیر آنها در نظر گرفته نمی‌شوند را به حداقل برسانند. شاید این مدل‌های پویا که به دنبال به حداقل رساندن پاداش هستند، غالبا به سمت یک تقریب تعادلی ناش تحت برخی شرایط همگرایی داشته باشند.

الگوریتم‌ها در تئوری یادگیری تقویتی می‌توانند به دو گروه مختلف تقسیم شوند که به توابع ارزشی که در طول آزمایش به روز می‌شوند بستگی دارد (فصل 15، 16، 17 و 21). در الگوریتم‌های یادگیری تقویتی که با نام بدون مدل شناخته می‌شود و در فصل 15 مورد بررسی قرار گرفتند، توابع ارزش برای تصمیم‌گیرنده بطور انحصاری با توجه به بازپرداختهای واقعی یا پاداشهای حاصل از اقدامات قبلی به روز می‌شوند. در مقابل، در الگوریتم‌های یادگیری تقویتی بر پایه مدل که در فصل 16 توضیح داده شد، رفتارها می‌توانند با انعطاف بیشتری طبق دانش تصمیم‌گیرنده نسبت به محیط خود تنظیم گردند.

ناحیه‌ای که در آن نوع از مدلها گسترش پیدا کرده‌اند در حوزه این است که تصمیم‌گیرنده در صورت انتخاب متفاوت چه پاداشی را دریافت کرده است. نتایج چنین اقدامات فرضی بعنوان نتایج ساختگی شناخته می‌شوند. مشابه با خطای پیش‌بینی پاداش سنتی، یادگیری تقویتی بدون مدل، تفاوت بین نتایج ساختگی و نتایج پیش‌بینی شده توابع ارزش فعلی بعنوان خطای پاداش ساختگی ارجاع داده می‌شوند. در یادگیری تقویتی بر اساس مدل مانند مدل جذب وزنی تجربه (EWA) کامرر و هو (1999)، توابع ارزش بطور مستقل و همزمان طبق دو خطای پیش‌بینی پاداش واقعی و ساختگی به روز می‌شوند. مطالعات تصویربرداری عصبی در انسان، سیگنالهایی را در ارتباط با خطاهای پیش‌بینی پاداش ساختگی در اجسام مخطط شناسایی کرده‌اند.

با این حال، نورون‌های دوپامین که خطاهای پیش‌بینی پاداش ساختگی را علاوه بر خطاهای پیش‌بینی پاداش واقعی رمزگذاری می‌کنند نیز هنوز ناشناخته است. فعالیت نورون‌های منفرد در ارتباط با نتایج ساختگی در نواحی قشر پیشانی شامل قشر مفصل قدامی و قشر پیشین حدقه‌ای شناسایی شده است.

### یادگیری تقویتی بدون مدل در طول بازی تطبیق پول در میمونها

در نسخه کلاسیک بازی تطبیق پول، هر کدام از دو بازیکن از دو گزینه جایگزین انتخاب خود را انجام می‌دهند و یکی از بازیکنان برنده خواهد شد و در صورتی برنده می‌شود که انتخابهای او مطابقت داشته باشد و در غیر اینصورت بازی را از دست می‌دهد. بازپرداخت به دیگر بازیکن در سمت مخالف است، بنابراین مجموع بازپرداخت دو بازیکن برابر صفر است. زمانی که دو بازیکن منطقی در بازی تطبیق پول شرکت می‌کنند، تعادل ناش برای بازیکنی است که دو هدف را با احتمال برابر و مستقل از آزمونهای موفق انتخاب کند. برای آزمایش این که چگونه میمونها استراتژیهای مطلوب تصمیم‌گیری را در بازیهای رقابتی از طریق تجربه تقریب می‌زدند، تعدادی از مطالعات رفتار انتخابی میمونها را در بازی تطبیق پول شبیه‌سازی شده با کامپیوتر بررسی کرده‌اند. در طول یکی از این آزمایشات نوروفیزیولوژی در میمونها، هر میمون بازی تطبیق پول را در مقابل رقیب کامپیوتری خود انجام داد. حیوان باید هر آزمون را با تثبیت یک مکعب زرد کوچک ارائه شده در مرکز صفحه کامپیوتر آغاز کند (دوره پیشین، شکل A26.2). بلافاصله پس از آن، دو صفحه سبز



در امتداد افقی ارائه شد و حیوان باید نگاه خود را زمانی که هدف ثابت مرکزی خاموش شد به سمت یکی از هدفها تغییر دهد. رقیب کامپیوتری همچنین یکی از این دو هدف را طبق الگوریتم توصیف شده در زیر انتخاب می‌کند (اگرچه این امر برای میمون نامرئی بود). حیوان تنها زمانی با پاداش روبرو می‌شد که همان هدف را که کامپیوتر انتخاب می‌کرد را انتخاب کند.

برای بررسی اینکه چگونه رفتار انتخابی حیوان تحت تاثیر استراتژی رقیب قرار می‌گیرد، استراتژی رقیب کامپیوتری به صورت سیستمی در مجموعه‌ای از آزمایشات توسط لی و همکاران (2004) دستکاری شد. در ابتدا، برای چند روز، رقیب کامپیوتری دو هدف را با احتمال یکسان بدون توجه به انتخابهای حیوان انتخاب کرد. این مورد تحت عنوان الگوریتم صفر در نظر گرفته شد، و مطابق با استراتژی تعادل ناش بازی تطبیق پول در تعادل ایستا است. در این مورد، سپس کامپیوتر این تعادل ایستا را بازی کرد، بازپرداخت قابل انتظار حیوان بدون توجه به انتخاب او ثابت شد. هر سه میمون با الگوریتم صفر آزمایش دادند و یک تمایل شدید را برای انتخاب یکی از دو هدف بطور مکرر نشان دادند.

در مرحله بعدی آزمایش، رقیب کامپیوتری مجموعه‌ای از آزمونهای آماری را برای انتخابهای میمون در جهت تعیین این که آیا تصمیمات میمون بطور تصادفی به دو هدف تقسیم می‌شود و اینکه آیا انتخابهای موفق بصورت آماری مستقل هستند به کار برد. اگر حیوان تمایلی را برای یک هدف یا عدم استقلال انتخابهای متوالی نشان دهد، کامپیوتر از این اطلاعات برای تطبیق انتخابهای آن استفاده می‌کند تا احتمالی که او بتواند هر مرحله را با پیروزی پشت سر بگذارد را به حداقل برساند. این رویکرد با پویایی بیشتر تحت عنوان الگوریتم یک در نظر گرفته شد. این الگوریتم تاریخچه پاداش حیوان را بررسی نکرد و بنابراین برای هر تمایلی که حیوان ممکن است برای انتخابهای آتی نشان دهد حساسیت نداشت. هنگامی که با الگوریتم یک آزمایش صورت گرفت، میمونها دو هدف را بصورت برابر بیشتر یا کمتر انتخاب کردند. علاوه بر این، انتخابهای موفق حیوان نسبتا مستقل بودند و در نتیجه نرخ پاداش کلی حیوان نزدیک به یک بود که ممکن است با دو بازیکن در تعادل ناش و به ارزش 0.5 به دست آید. جالب توجه است که حیوانات بیشتر تمایل داشتند که هدف مشابه را در آزمون بعدی انتخاب کنند و این در صورتی رخ می‌داد که انتخاب قبلی آنها با پاداش همراه بوده باشد (برنده ماندن) و تغییر به سمت دیگر هدف با شکست همراه است (شکست در صورت تغییر).

چنین استراتژیهای برنده و شکست در طول دوره الگوریتم یک مجازات نشدند، زیرا اطلاعات در مورد تاریخچه پاداش حیوان توسط رقیب کامپیوتری استفاده نشد. هر سه حیوان انتخابهای خود را طبق استراتژی فوق در بیش از 50 درصد آزمایشها انجام دادند.

در مرحله نهایی این آزمایشات رفتاری بر روی فعالیت تطبیق پول، رقیب کامپیوتری (الگوریتم 2) از تمایلات در انتخاب حیوان و تاریخچه پاداش استفاده کرد که شامل تمایل به استفاده از استراتژی WSLs بود. زمانیکه این مورد رخ داد، حیوانات تمایل کمتری برای کسب پاداش در صورت استفاده از استراتژی WSLs در بیش از 50 درصد آزمونها داشتند. همانطور که انتظار میرفت، این امر احتمال استراتژی WSLs را بطور قابل توجهی در حیوانات کاهش داد. با این حال، استراتژی WSLs کماکان بیش از درصد در حیوانات مورد استفاده قرار می گرفت، که نشان دهنده این است که حیوانات هنوز به الگوریتم یادگیری تقویتی برای تقریب زدن استراتژی تعادلی ناش در طول فعالیت تطبیق پول اتکا دارند.

در مدل‌های یادگیری تقویتی، احتمال انتخاب یک اقدام خاص معمولاً توسط بیشینه هموار یا تبدیل لجستیکی توابع ارزش برای تمامی اقدامات داده می شود. زمانیکه تنها دو انتخاب وجود داشته باشد، این معادله به معادله زیر کاهش پیدا می کند.

$$P_t(\text{right}) = \frac{e^{\beta Q_t(\text{right})}}{e^{\beta Q_t(\text{right})} + e^{\beta Q_t(\text{left})}} \quad (26.1)$$

جاییکه  $P_t$  (راست) به احتمال انتخاب حرکت به سمت راست در آزمون  $t$  دلالت دارد و  $Q(x)$  نیز تابع ارزش انتخاب اقدام  $x$  (که  $x$  معادل راست یا چپ است) است. پارامتر  $\beta$  نیز تصمیم تصادفی را اتفاقی تصمیم گیرنده را تعیین می کند. هنگامی که مقدار  $\beta$  برابر با صفر باشد، انتخابها کاملاً تصادفی هستند و ارتباطی به توابع ارزش ندارند و زمانیکه مقدار  $\beta$  افزایش پیدا می کند قطعی تر می شوند. در الگوریتم یادگیری تقویتی بدون مدل استاندارد، توابع ارزش برای اقدامات انتخاب شده توسط تصمیم گیرنده (که بصورت عبارات  $Q$  فهرست می شوند) مطابق با معادله زیر تنظیم می گردند:

$$Q_{t+1}(x) = Q_t(x) + \alpha \{\text{reward}_t - Q_t(x)\} \quad (26.2)$$

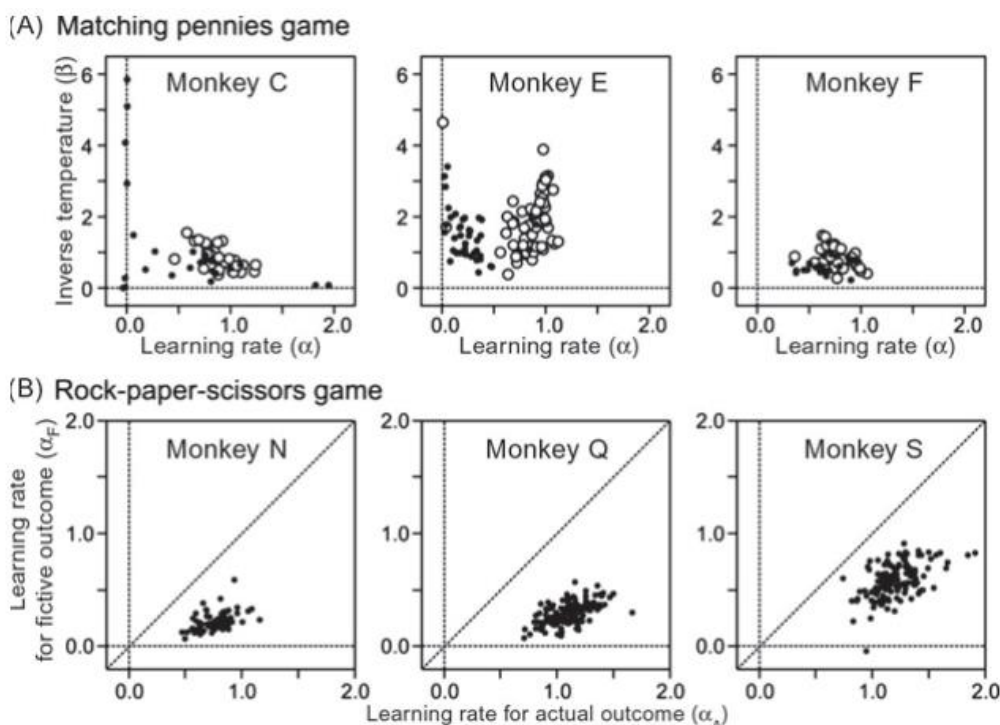
جاییکه reward<sub>t</sub> نشان دهنده پاداش دریافت شده توسط تصمیم‌گیرنده است (1 و صفر به ترتیب برای آزمونهای پاداش داده شده و نشده) و آلفا نیز نرخ یادگیری می‌باشد.

لی و همکاران (2004) بصورت تجربی دریافته‌اند که انتخابهای میمونها در طول بازی تطبیق پول نسبتاً تصادفی بود. انتخابهای حیوانات همچنین به خوبی توسط مدل یادگیری تقویتی بدون مدل توجیه شد. علاوه بر این، این واقعیت که احتمال استفاده از استراتژی WLS در برابر رقیب کامپیوتری که در حال استفاده از الگوریتم 2 کاهش پیدا می‌کند، نشان می‌دهد که این امر ممکن است به دلیل نرخ یادگیری کوچکتر باشد. در عوض، این امر همچنین می‌تواند در نتیجه مقدار بتای کوچکتر باشد، که باعث تصادفی‌تر شدن انتخابهای حیوان می‌شود. پارامترهای تخمینی برای داده‌های رفتاری حیوان نشان می‌دهد که تغییرات در انتخابهای حیوان در ارتباط با الگوریتم‌های مختلف رقیب کامپیوتری بطور عمده به دلیل تغییرات در نرخ یادگیری بود (شکل 26.4). این نتایج نشان می‌دهد که زمانیکه در طول بازی رقابتی با رقیب کامپیوتری روبرو شدند، حیوانات انتخابهای تصادفی‌تری را داشتند که شاید با کاهش نرخ یادگیری همراه بود. علاوه بر این، آنها نمونه خوبی از یادگیری متا را فراهم می‌کنند که در آن پارامترهای مدل یادگیری مانند نرخ یادگیری بهینه‌سازی می‌شوند.

### یادگیری ترکیبی در طول بازی سنگ-کاغذ-قیچی در میمونها

در یادگیری تقویتی بدون مدل که در بالا توصیف شد، تنها تابع ارزش برای اقدام انتخاب شده توسط تصمیم‌گیرنده در یک آزمون داده شده با توجه به نتیجه آن اقدام به روز می‌شود. در مقابل، نتایج مطالعات بر روی بازیهای آزمایشی در انسان نشان می‌دهد که افراد می‌توانند توابع ارزش را برای اقدامات انتخاب نشده مطابق با نتایج قابل فهمی که اقدامات جایگزین ممکن است داشته باشند تنظیم کنند، ویژگی الگوریتم EWA که در بالا توصیف شد. اخیراً یافت شده است که میمونها نیز می‌توانند استراتژی‌های خود را مطابق با نتایج قابل فهم اقدامات انتخاب نشده در طول بازی سنگ-کاغذ-قیچی تنظیم کنند. با این وجود، مطابق با نتایج حاصل از مطالعات در انسانها، انتخابهای میمونها بیشتر تحت

تاثیر نتایج واقعی حاصل از اقدامات انتخاب شده توسط حیوان بود تا اینکه تحت تاثیر نتایج قابل فهم اقدامات انتخاب نشده باشد.



شکل 26.4. عملکرد رفتاری میمونها در طول بازی تطبیق پول (آ) و سنگ-کاغذ-قیچی (ب). (آ) دماهای معکوس (بتا) و نرخ یادگیری (آلفا) که از جلساتی که در آن رقیب کامپیوتری هدف خود را با استفاده از الگوریتم 1 (حلقه-های خالی) و 2 (نقاط) در طول فعالیت تطبیق پول نشان داده شده برای سه حیوان مختلف انتخاب کرد تخمین زده شد. (ب) نرخهای یادگیری برای نتایج واقعی و قابل فهم که با استفاده از مدل یادگیری ترکیبی برای تناسب رفتارها در طول فعالیت سنگ-کاغذ-قیچی تخمین زده شد و برای سه حیوان مختلف نشان داده شده است.

برای توضیح این امر، آبه و لی (2011) میمونها را برای بازی سنگ-کاغذ-قیچی آموزش دادند (شکل 26.1). میمونها ابتدا یک هدف مرکزی را در ابتدای هر آزمون ثابت کردند (شکل 26.2). بعد از یک تاخیر کوتاه، سه صفحه سبز بعنوان اهداف انتخابی ارائه شد و حیوان آزاد بود که نگاهش را به سمت یکی از این سه هدف در زمان ناپدید شدن هدف مرکزی تغییر دهد. هر کدام از این سه هدف بصورت سنگ، کاغذ یا قیچی طراحی شدند و حیوان در نتیجه این انتخاب با پاداش روبرو می شد و مقداری از آب به حیوان داده می شد که با ماتریس بازپرداخت بازی در رقابت با رقیب

کامپیوتری تعیین می‌شد. برای مثال، اگر حیوان هدف سنگ را انتخاب میکرد و کامپیوتر نیز هدف کاغذ را بر می‌گزید، به حیوان هیچ پاداشی داده نمی‌شد. زمانیکه نتیجه برابر با کراوات میشد، حیوان با یک قطره آب تشویق می‌شد. زمانیکه حیوان با انتخاب سنگ، کاغذ و قیچی برنده می‌شد به ترتیب دو، سه و چهار قطره آب را پاداش می‌گرفت. بازپرداخت برای برنده شدن آزمون متنوع بود بطوریکه اثرات عصبی و نوروفیزیولوژیکی نتایج درک شده می‌تواند بصورت کمی سنجیده شود (شکل 26.1). در طول این آزمایش، لازم نبود که حیوانات قواعد بازی سنگ-کاغذ-قیچی را به یاد داشته باشند، زیرا بازپرداخت حاصل از سه انتخاب توسط انتخاب رقیب کامپیوتری و با استفاده از رنگهای محرک بازخورد تعیین می‌شد (شکل 26.2).

برای اینکه چگونه انتخابهای حیوان در طول این بازی تحت تاثیر نتایج قابل فهم قرار می‌گرفت، داده‌های رفتاری حاصل در طول این آزمایش با چندین مدل یادگیری مختلف مورد تجزیه و تحلیل قرار گرفت (آبه و لی، 2011). این شامل مدل یادگیری تقویتی بدون مدل (مانند موردی که در بالا توصیف شد) و نیز مدل یادگیری باور بود. در مدل یادگیری باور، بازیکنان اعتقادات و باورهای خود را در مورد استراتژیهای دیگر بازیکنان بعد از هر آزمون به روز می‌کنند و انتظار می‌رود که انتخابهای آنها بهترین نتایج را با توجه به این باورها داشته باشد. این مدل برای انتخابهای حیوان در طول بازی سنگ-کاغذ-قیچی با به روزرسانی توابع ارزش برای هر سه گزینه مطابق با نتایج تعیین شده توسط انتخاب رقیب کامپیوتری اعمال شد. برای مثال، زمانیکه کامپیوتر هدف سنگ را انتخاب کرد، نتیجه برای حیوان در صورت انتخاب سنگ، کاغذ و قیچی به ترتیب به صورت 1، 2 و صفر است، و این مقادیر بعنوان پاداشهای قابل فهم برای به روزرسانی توابع ارزش برای سنگ، کاغذ و قیچی مورد استفاده قرار گرفت. در نهایت، در یک مدل یادگیری ترکیبی، توابع ارزش برای اقدامات انتخاب شده و انتخاب نشده با استفاده از دو نرخ یادگیری مجزا به روز شدند که به صورت زیر است:

$$\begin{aligned}
 Q_{t+1}(x) &= Q_t(x) + \alpha_A \{ \text{actual\_reward}_t - Q_t(x) \}, \\
 &\text{if } x \text{ was chosen, and} \\
 Q_{t+1}(x) &= Q_t(x) + \alpha_H \{ \text{fictive\_reward}_t(x) - Q_t(x) \}, \\
 &\text{if } x \text{ was not chosen}
 \end{aligned}
 \tag{26.3}$$

جاییکه  $fictive-reward_t(x)$  به پاداش قابل فهمی که می‌تواند از انتخاب  $x$  در آزمون  $t$  حاصل شود بر می‌گردد. علاوه بر این،  $\alpha_A$  و  $\alpha_F$  نیز به ترتیب به نرخ یادگیری برای نتایج واقعی و قابل فهم اختصاص دارند. نتایج حاصل از تحلیلها نشان داد که مدل یادگیری ترکیبی انتخابهای میمونها را در طول فعالیت سنگ، کاغذ، قیچی بهتر از مدل یادگیری تقویتی بدون مدل و مدل یادگیری باور توجیه می‌کند. علاوه بر این، نرخهای یادگیری برای نتیجه قابل فهم نیز همیشه کمتر از موارد مربوط به نتایج واقعی بود (شکل 26.4) که نشان دهنده این است که نتایج واقعی تاثیرات قویتری را بر انتخابهای متوالی حیوان دارد.

### مکانیک قشری یادگیری تقویتی در طول بازیهای تکراری

#### فعایت عصبی مربوط به مقادیر و انتخابها

یکی از نواحی اولیه فرضیه که در ارائه ارزش اهداف بصری در روندی که می‌توانند برای انتخاب اقدامات استراتژیک استفاده شوند عبارتست از ناحیه قشر پس سری (LIP). این ناحیه برای مطالعه انتخاب شد زیرا در انتهای جریان پردازش بصری قرار دارد و خروجی‌های آن بر نواحی مغز از جمله برنامه‌ریزی و حرکات چشم تاثیر می‌گذارد. کار قبلی نشان داده بود که این فعالیت در این ناحیه ممکن است اهمیت اهداف بصری را در روندی که بتواند به منابع قابل توجه تخصیص داده شود را رمزگذاری کند و یا بین اهداف حرکتی چشم انتخابی را صورت دهد. یک مطالعه پیشگام که توسط پلات و گلیمچر (1999) انجام شد نشان داد که متغیرهای اقتصادی مهم مانند احتمال و مقدار پاداش بر روی نرخ کارکرد نوروهای LIP تاثیر می‌گذارد و یک چارچوب نظری تصمیم‌گیری را برای مطالعه نقش نواحی مغزی در تحولات حسی-حرکتی فراهم می‌کند.

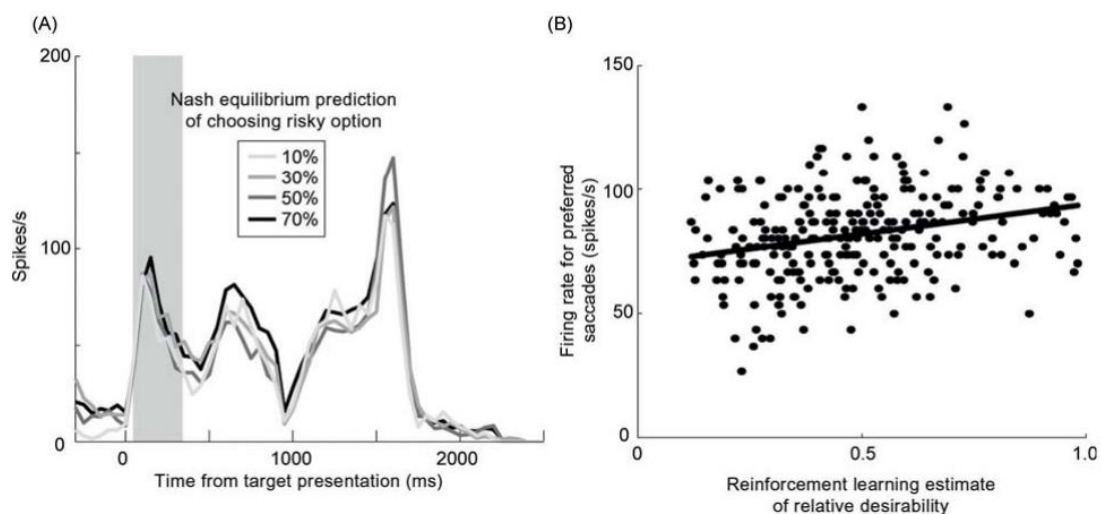
با توجه به اینکه ناحیه LIP در ارتباط بین پردازش حسی و حرکتی قرار دارد و تحت تاثیر متغیرهای اقتصادی است، دوریس و گلیمچر (2004) فرض کردند که این ناحیه می‌تواند نقش مهمی را در ارائه ارزش ذهنی اهداف انتخابی (یک همبستگی عصبی از اهداف اقتصادی مانند ابزار مورد انتظار) در طول بازیهای رقابتی بازی کند. در آزمایش آنها، میمونها در مقابل یک حریف کامپیوتری در طول بازی بازرسی با استراتژی مخلوط به رقابت پرداختند (شکل 26.2). از دیدگاه

میمون، هدف مخالف میدان پاسخ باعث دریافت مقدار خاصی از آب در صورت انتخاب می‌شد. هدف موجود در ناحیه پاسخ خطرناک بود و می‌توانست دو برابر آب یا هیچ مقداری از آب را به دنبال داشته باشد. ماتریس بازپرداخت بطور تجربی در سراسر بلوکهای آزمایش تعیین شد بطوریکه راه حل تعادل ناش در استراتژی مخلوط برای میمون در محدوده انتخاب هدف در مرکز ناحیه پاسخ نورون از 10 تا 90 درصد زمان بود. این تعادل با دستکاری هزینه بازرسی رقیب کامپیوتری دستکاری شد (شکل 26.2، متغیر ا).

در صورتی که حتی ا کم نیز باشد، تعادل طوری تغییر می‌کند که گزینه خطرناک را ریسکی به ندرت انتخاب شود، در حالیکه اگر این مقدار زیاد باشد، تعادل طوری تغییر می‌کند که گزینه همراه با ریسک بطور مکرر انتخاب شود. احتمال رقیب کامپیوتری در بازرسی برابر با 50 درصد در تعادل مستقل از مقدار ا باقی می‌ماند، ویژگی اصلی تئوری بازی. دوریس و گلیمچر (2004) دریافتند که انسانها و میمونها هر دو به فرکانسهای تعادلی پیش‌بینی شده در هنگام بازی با رقیب کامپیوتری نزدیک شدند، هرچند آنها تمایل داشتند که گزینه ریسکی را کمی بیشتر در زمانیکه هزینه بازرسی پایین بود انتخاب کنند. آنها این استدلال را داشتند که اگر LIP، احتمال حرکت را رمزگذاری کند، فعالسازی آن ممکن است در طول بلوکهای آزمایش تغییر کند چرا که احتمالات حرکت تغییر پیدا کرد. با این حال اگر LIP، مقدار و ارزش ذهنی اهداف را رمزگذاری کند، فعالسازی آن باید نسبتاً ثابت باقی بماند زیرا نظریه بازی نشان می‌دهد که این مقدار در تعادل استراتژی مخلوط که مستقل از احتمالات حرکت است ثابت باقی می‌ماند. تفسیر آخر یک ویژگی مهم در مفهوم تعادل ناش ارائه شده در فصل 2 است، و از این واقعیت پیروی می‌کند که نتیجه‌گیری نظری توسط ناش این است که ارزش ذاتی (یا ابزار مورد انتظار) باید بطور متوسط در بین گزینه‌های مخلوط در بازی تعادل استراتژی مخلوط برابر باشد. در واقع، فعالیت LIP با ارزش ذهنی محرک انتخابی شکل می‌گیرد؛ نرخهای تبادل نورون‌ها با تغییر ارزش تحت شرایط گزینه اجباری متغیر بود و در سراسر تعادل رفتاری ایجاد شده در طول شرایط استراتژی مخلوط ثابت ماند (شکل 26.5).

اگرچه مفهوم تعادل ناش بر این اصل استوار است که می‌تواند بدون داشتن انگیزه برای تغییر استراتژی کلی شخص در تعادل رفتاری باشد، اما مشخص نمی‌کند که در سطح آزمون به آزمون چه مفهومی می‌تواند داشته باشد. سیگنالهای

دقیق حاصل از ثبت نورون‌های منفرد به ما این اجازه را می‌دهد که بررسی کنیم آیا LIP با ارزش ذهنی در سطح آزمون به آزمون بعنوان تابعی از انتخابی که شخص انجام می‌دهد همبستگی دارد. برای تخمین ارزش ذهنی بر اساس آزمون به آزمون، دوریس و گلیمچر (2004) یک الگوریتم یادگیری تقویتی بدون مدل ساده را مشابه موارد توصیف شده در بالا و فصل 15 بهینه‌سازی کردند. آنها مدل را متناسب با الگوی انتخاب رفتاری میمون با استفاده از روشهای احتمال بیشینه به منظور امتحان و پیش‌بینی الگوی میمون در انتخاب در سطح آزمون به آزمون قرار دادند. بطور خلاصه، آنها ارزش ذهنی هر گزینه را اینگونه فرض کردند که در صورت دریافت پاداش برای انتخاب گزینه ریسکی افزایش پیدا می‌کند و اگر پاداشی داده نشود کاهش با کاهش روبرو می‌شود. تنها پارامتر آزاد عبارت بود از نرخ یادگیری که در آن مقدار و ارزش بر اساس این اطلاعات پاداش به روز می‌شد. ماهیت تکراری این الگوریتم یادگیری تقویتی منجر به تخمین ارزش ذهنی مشتق شده از همه انتخابهای قبلی شد. نویسندگان دریافتند که نوسانات آزمون به آزمون در فعالیت LIP با این تخمین رفتاری ارزش ذهنی تغییر پیدا می‌کند (شکل 26.5).



شکل 26.5. رمزگذاری ارزش ذهنی اهداف بصری در ناحیه LIP. (آ) فعالیت نورون در طول بازی بازرسی در حالت استراتژی مخلوط. با وجود تغییرات در احتمال پاسخهای ترجیح داده شده، فعالیت LIP نسبتاً ثابت باقی می‌ماند که مطابق با تعادل کلی در ارزش ذهنی در تعادل استراتژی مخلوط است. (ب) تغییر آزمون به آزمون در فعالیت در طول دوره بصری بطور قابل ملاحظه‌ای با تخمین رفتاری ارزش ذهنی همبستگی داشت. برگرفته از مطالعه دوریس و

گلیمچر (2004).



علاوه بر ناحیه LIP که ارزش هدف و انتخابهای حرکتی را منعکس می‌کند، در بسیاری از نواحی دیگر مغز نیز از جمله قشر پیشانی شناسایی شده است. فعالیت نورون‌ها در هر کدام از نواحی در ارتباط با توابع ارزش برای اقدامات خاص یا تحولات آنها است. لی و همکارانش این امر را در مجموعه‌ای از مطالعات که در آنها فعالیت از نورونهای منفرد در قشر پیشانی، قشر پشتی قدامی و LIP ثبت شد توضیح دادند. نتایج این مطالعات نشان داد که بلافاصله قبل از اینکه حیوان هدفش را انتخاب کند (در طول دوره تاخیر، شکل 26.2)، نورونها در همه این نواحی نه تنها انتخاب بعدی را رمزگذاری می‌کنند، بلکه همچنین مجموع توابع ارزش برای دو اقدام مختلف و اختلاف آنها را نیز رمزگذاری می‌کنند. این حالت با استفاده از مدل رگرسیون زیر برای بررسی نرخ عمل عصبی نشان داده شد:

$$S_t = b_0 + b_1 C_t + b_2 \{Q_t(\text{right}) - Q_t(\text{left})\} + b_3 \{Q_t(\text{right}) + Q_t(\text{left})\} \quad (26.4)$$

جاییکه  $S_t$  عبارتست از نرخ عمل نورون داده شده در طول دوره تاخیر در آزمون  $t$ ،  $C_t$  نیز انتخاب حیوان است (اگر حیوان هدف راست را انتخاب کند این مقدار برابر 1 است و در غیر این صورت برابر صفر می‌باشد)، و تابع ارزش برای هدف  $x$  در آزمون  $t$  نیز برابر  $Q_t(x)$  است که از یادگیری تقویتی بدون مدل که در بالا توصیف شد تخمین زده شد. تفاوت در توابع ارزش استفاده شده در این مدل ممکن است توسط حیوان برای تعیین انتخاب خود استفاده شود، در حالیکه مجموع آنها ممکن است در ارتباط با تابع حالت ارزش باشد. تابع حالت ارزش با میانگین توابع ارزش وزنی توسط احتمال هر اقدام ارتباط دارد و بنابراین خوبی کلی گزینه‌های روبرو در زمان داده شده را نشان می‌دهد. برای مثال، در طول بازی تطبیق پول، هر دو اقدام با احتمال تقریباً مساوی انتخاب می‌شود و متوسط توابع ارزش تخمین خوبی از تابع حالت ارزش است. نتایج این تحلیل نشان داد که سیگنالهای مرتبط با مجموع توابع ارزش در مغز و در سطح نورونهای منفرد در حال گسترش هستند.

علاوه بر این، نسبت قابل توجهی از نورونها در  $Dipfc$  و  $LIP$  (اما نه در  $ACC$ ) فعالیت خود را طبق اختلاف در توابع ارزش مدوله کردند. این نتایج نشان می‌دهد که شبکه قشری متشکل از نواحی پیشانی و جداری ممکن است برای انتخاب عمل بر پایه ارزش در طول بازیهای رقابتی تکراری ارزشمند باشند. همچنین به نظر می‌رسد که سیگنالهای بر پایه ارزش که در این نواحی مغز در طول بازی تطبیق پول مشاهده می‌شوند به احتمال زیاد منجر به یادگیری تقویتی

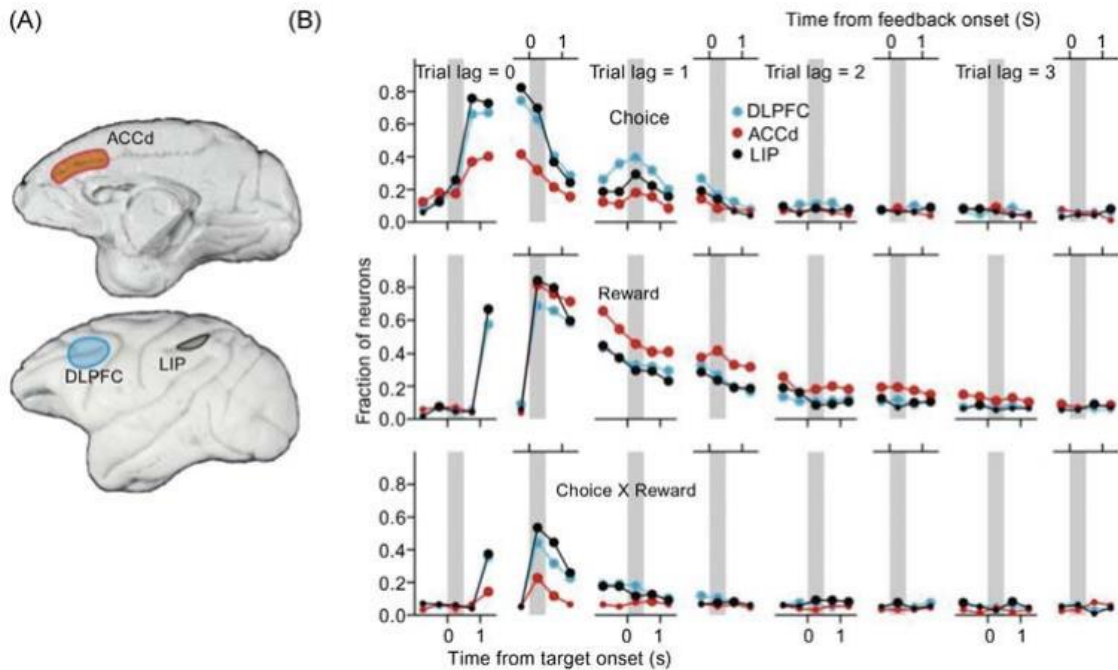
در قشر غیر اجتماعی می‌شوند، که در آن انتخابهای شخص بخوبی توسط الگوریتمهای یادگیری تقویتی بدون مدل توصیف می‌شود.

### فعالیت عصبی مرتبط با تاریخچه انتخاب و پاداش

نتایج توصیف شده در بخش قبلی نشان می‌دهد که نورونها در نواحی قشری مختلف مانند DIPFC و LIP ممکن است نقش مهمی را در ادغام سیگنالهای مرتبط با انتخابهای قبلی حیوان و نتایج آنها برای به روزرسانی توابع ارزش بازی کنند. برای آزمایش مستقیم این کار، لی و همکارانش مدل رگرسیون زیر را به کار بردند که شامل انتخابهای قبلی حیوان و رقیب کامپیوتری و نیز نتایج انتخاب حیوان است:

$$S_t = B[1 \ u_t \ u_{t-1} \ u_{t-2} \ u_{t-3}]' \quad (26.5)$$

جاییکه  $u_t$  عبارتست از بردار خطی متشکل از سه متغیر ساختگی مرتبط با انتخاب حیوان (صفر یا 1 به ترتیب برای انتخابهای چپ و راست)، انتخاب کامپیوتر (به همان شیوه انتخاب حیوان کدگذاری شد) و پاداش (1 برای آزمونهای پاداش داده شده و صفر برای آزمونهایی که با پاداش همراه نبوده است) در آزمون  $t$ ، و  $B$  نیز بردار 13 ضریب رگرسیون است. بنابراین، ضرایب رگرسیون در این مدل بطور کمی مشخص می‌کنند که چقدر فعالیت نورون داده شده توسط انتخابهای فعلی و قبلی حیوان و نتایج آنها و نیز انتخابهای رقیب کامپیوتری مدول شده است. این تحلیل بطور مجزا برای نرخهای عمل اندازه‌گیری شده با مجموعه‌ای از جاهای ذخیره‌سازی غیرهمپوشان 0.5 ثانیه‌ای که نسبت به زمان آغاز هدف یا بازخورد تعریف شدند مورد اجرا قرار گرفت. نتایج نشان داد که بسیاری از نورون‌ها در DIPFC و LIP سیگنالهای مرتبط با انتخاب حیوان و نتایج آن و نیز انتخاب رقیب کامپیوتری را نه تنها در آزمون فعلی، بلکه همچنین در چند آزمون قبلی را کدگذاری کردند (شکل 26.6). فعالیت مرتبط با انتخابهای قبلی رقیب کامپیوتری ممکن است در ارتباط با توابع ارزش برای انتخابهای جایگزین باشد، زیرا در طول بازی تطبیق پول به حیوان برای انتخاب هدف مشابه با رقیب کامپیوتری پاداش داده می‌شد.



شکل 26.6. نواحی قشری (آ) و خلاصه‌ای از فعالیت عصبی (ب) که در طول فعالیت تطبیق پول مورد بررسی قرار گرفت. (آ) موقعیت‌های قشر قدامی جانبی پس سری (ACCd)، قشر خلفی جانبی پیشانی (DLPFC) و LIP. (ب) دوره زمانی سیگنال‌های مرتبط با انتخاب حیوان (بالا)، پاداش (وسط) و پیوندهای آنها (پایین). هر ردیف نشان دهنده نسبت نورون‌ها در هر ناحیه قشری است که بطور قابل ملاحظه‌ای فعالیت آنها را مطابق با انتخاب حیوان، پاداش یا اتصالات آنها (یا انتخاب کامپیوتر) در آزمون‌های فعلی (trial lag=0) و سه آزمون قبلی (trial lag=1,2,3) مدوله کرد. نماد بزرگ نشان می‌دهد که اثرات یافت شده در بیشتر نورون‌ها بیشتر از حد انتظار بود.

سیگنال‌های مرتبط با انتخاب‌های قبلی حیوان ممکن است بصورت سیگنال‌های حافظه موقت که تاریخچه انتخاب حیوان را رمزگذاری می‌کند عمل کنند. در نظریه یادگیری تقویتی، پاداش‌های تاخیری که مختص اقدامات قبلی هستند بعنوان مسئله تخصیص اعتبار موقت ارجاع داده می‌شود و سیگنال‌های حافظه مربوط به انتخاب‌های قبلی حیوان نیز غالباً بعنوان اثر شایستگی ارجاع داده می‌شود، که می‌تواند این مسئله که در فصل 15 مورد بحث قرار گرفت را حل کند. بنابراین، فعالیت عصبی مرتبط با انتخاب‌های قبلی حیوان که در هر دو LIP و diPFC دیده شدند، ممکن است مربوط به اثرات شایستگی در مدل‌های یادگیری تفاوت زمانی باشند.

فعالیت مربوط به تاریخچه پاداش حیوان نیز در قشر پیشانی و قشر جداری خلفی دیده شد. سیگنالهای مربوط به تاریخچه پاداش بطور خاصی در ACC قوی بودند که مطابق با این ایده است که قشر میانی پیشانی نقش مهمی را در نظارت بر نتایج اقدامات مختلف ایفا می‌کند. فعالیت مربوط به تاریخچه پاداش حیوان ممکن است نقش مهمی را در محاسبه نرخ متوسط پاداش و نحوه انحراف یک پاداش خاص از پاداش مورد انتظار و خطای پیش‌بینی پاداش بازی می‌کند. جالب توجه است که، سیگنالهای مربوط به انتخاب حیوان و تاریخچه پاداش یافت شده در این نواحی قشری مختلف ناهمگن بودند و ثابتهای زمانی آنها به خوبی توسط تابع قدرت توصیف شد، که نشان دهنده این است که ثابتهای زمانی برای سیگنالهای مربوط به انتخابهای قبلی و نتایج ممکن است در تعدادی از نورون‌ها نسبتاً طولانی باشد. این امکان را ایجاد می‌کند که نورون‌ها در این نواحی قشری مختلف، یک مخزن از ثابتهای زمانی را که می‌توانند مطابق با مقیاس زمانی بهینه برای فعالیت رفتاری خاص انتخاب شوند را فراهم سازد (فصل 23).

### فعالیت عصبی مرتبط با نتایج ساختگی (فکری)

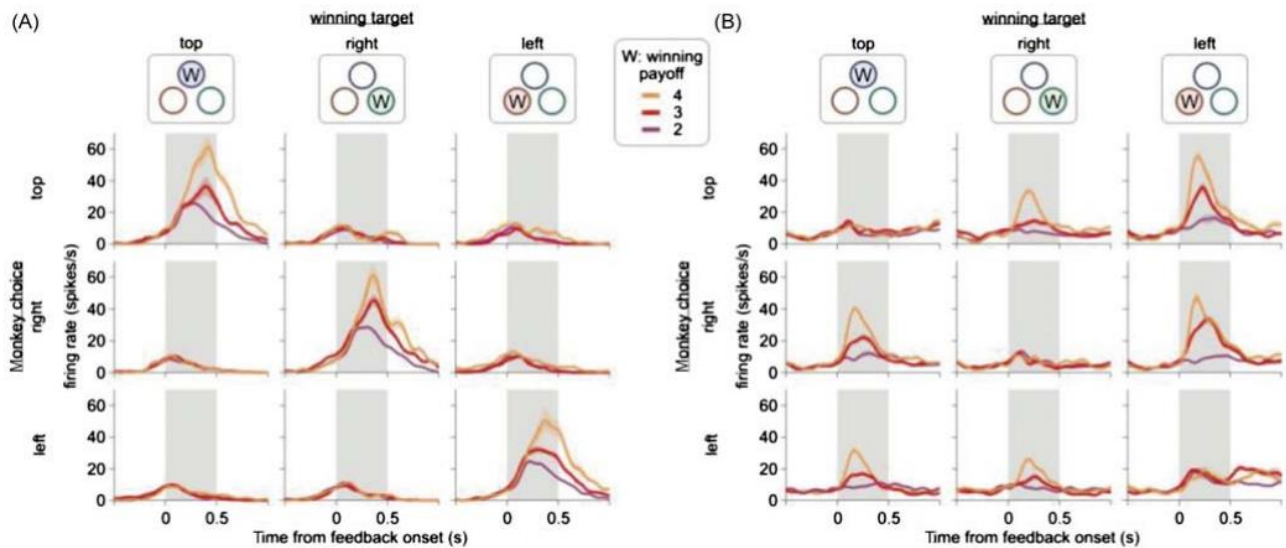
تحلیل‌های داده‌های رفتاری حاصل از آزمایش سنگ-کاغذ-چیچی که در بالا توصیف شد نشان داده است که نه تنها نتایج واقعی اقدامات انتخابی توسط حیوان، بلکه همچنین نتایج فکری و ساختگی اقدامات انتخاب نشده نیز بر انتخابهای بعدی حیوان تاثیر می‌گذارد. برای تعیین اینکه آیا قشر پیشانی در ترکیب هر دو نتایج واقعی و ساختگی در توابع ارزش مختلف دخالت دارد، فعالیت نورون‌های منفرد در diPFC و قشر پیشین حدقه‌ای در میمون‌هایی که در حال بازی سنگ-کاغذ-چیچی بودند ثبت شد. مطابق با یافته‌های حاصل از مطالعات قبلی، نتایج این مطالعه نشان داد که نورون‌ها در diPFC و OFC غالباً نتایج واقعی را از انتخابهای حیوان رمزگذاری می‌کنند. فعالیت مربوط به نتیجه واقعی غالباً در طول دوره بازخورد که در آن اطلاعات در مورد نتیجه واقعی حاصل از هدف انتخابی و نتایج ساختگی حاصل از هدف انتخاب نشده برای حیوان مشخص شد ملاحظه گردید (شکل 26.2). برای مثال، نورون OFC نشان داده شده در شکل 26.7، فعالیت خود را با افزایش دادن پاداش کسب شده توسط حیوان در طول دوره بازخورد آزمونهایی که با پیروزی همراه بودند افزایش داد. نورون‌ها در هر دو diPFC و OFC نیز نتایج حاصل از اقدامات خاص را رمزگذاری

کردند. برای مثال، برخی از نورون‌ها فعالیت خود را مطابق با نتایج حاصل از انتخاب سنگ تغییر دادند، در حالیکه دیگر نورون‌ها فعالیت خود را مطابق با نتایج حاصل از انتخاب کاغذ مدوله کردند. این گرایش در diPFC قویتر از OFC بود، که نشان دهنده این است که ممکن است diPFC نقش مهمتری را در به روز کردن توابع ارزش داشته باشد. مهمتر از همه، نورون‌هایی که نتایج ساختگی را رمزگذاری می‌کنند نیز در هر دو diPFC و OFC دیده شدند. نورون OFC نشان داده شده در شکل 26.7 فعالیت خود را بصورت جزئی تنها در طول دوره بازخورد آزمونهای پیروز تغییر داد، اما فعالیت خود را به صورت سیستماتیک مطابق با افزایش پاداش ساختگی که حیوان می‌تواند کسب کند در آزمونهایی که با شکست همراه بود با انتخاب یکی از اهداف باقیمانده افزایش داد. برای برخی از نورون‌ها در هر دو diPFC و OFC، فعالیت مربوط به پاداش ساختگی حاصل از اهداف پیروز غیر انتخاب شده طبق موقعیت هدف پیروز تغییر پیدا کرد، و این گرایش در diPFC به مراتب قویتر از OFC بود. این نتایج نشان می‌دهد که diPFC و OFC نقش مهمی را در رمزگذاری نتایج واقعی و نتایج ساختگی حاصل از اقدامات انتخاب نشده بازی می‌کند، و ممکن است از این سیگنالها برای به روزرسانی توابع ارزش برای هر دو اقدامات انتخابی و انتخاب نشده که قبلاً در یادگیری تقویتی بر پایه مدل توصیف شد استفاده کنند.

### انتخاب پاسخ توسط نواحی چشمی پیشانی و برجستگی فوقانی

نواحی جداری نشان داده شده در بالا (مانند diPFC، ACC، OFC و LIP) به نظر می‌رسد که آمار مهمی برای محدوده تصمیم‌گیری اجتماعی از تاریخچه قبلی انتخابها و نتایج آنها تا ارزیابی انتخابها و نتایج آنها و تا توابع ارزش‌گذاری و حتی تا دانش در مورد اینکه چه چیزی می‌تواند بوده باشد ارائه می‌دهد. با این حال بعید است که هر یک از این نواحی در نهایت پاسخ انتخابی را انتخاب و اجرا می‌کنند. این بدان معناست که مشکل در راه‌اندازی حرکت با تحریک میکرو در این نواحی، همبستگی‌های ضعیف فعالیت با زمانهای واکنش حرکتی بر روی ایجاد حرکت تاثیر می‌گذارند که در نتیجه آسیب این نواحی است. برجستگی زبرین (SC) و یکی از منابع ورودیهای قشری به نام نواحی چشمی پیشانی (FEF) همگی در در انتخاب حرکات و ایجاد دستورات حرکتی چشم دخیل هستند. حرکات چشم با تحریک میکرو

الکتریکی در این نواحی در جریانهای پایین تحریک می‌شوند، الگوهای فعالیت پیش‌بینی کننده هر دو است و این اتفاق زمانی رخ می‌دهد که حرکت چشم رخ می‌دهد. هر دو برای مدارهای مغزی که بطور مستقیم نیروهای عضلانی و حرکتیهای چشمی را کنترل می‌کنند، اگر این دو ساختار تخریب شوند دیگر قادر به عمل نخواهند بود. در این بخش ما در مورد چگونگی فعالیت در FEF و SC برای انتخاب پاسخ حرکتی در طول بازی با استراتژی مخلوط و تطبیق پول بحث می‌کنیم.



شکل 26.7. نمونه‌هایی از نوروهای ثبت شده در قشر پیشین حلقه‌ای میمون‌ها که نتایج (آ) واقعی و (ب)

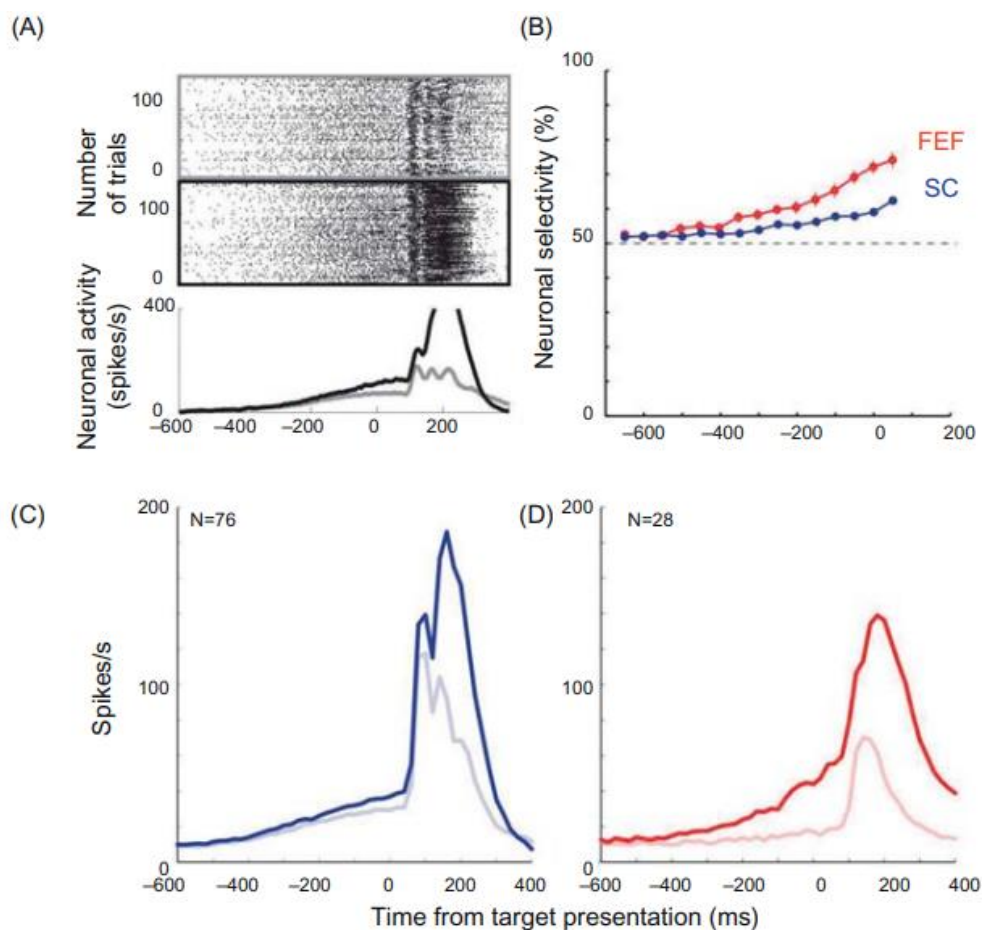
ساختگی را در طول فعالیت سنگ-کاغذ-قیچی رمزگذاری کردند. فعالیت هر نورون بر حسب انتخاب حیوان (ردیف-ها)، انتخاب کامپیوتر (ستونها) و بازپرداخت حاصل از هدفی که منجر به پیروزی شده است (رنگهای مختلف) رسم گردید. پس زمینه خاکستری، مربوط به دوره بازخورد 0.5 ثانیه است.

آزمایش تطبیق پول، پیچیده‌ترین رقیب کامپیوتری را مدیون لی و همکاران (2004) است، الگوریتم سطح 2 که در بالا (شکل 26.2) با دو استثنای مهم نشان داده شد. اول، در طول هر جلسه آزمایشی، موقعیتهای اهداف انتخابی با ناحیه پاسخ نورون تحت مطالعه متناسب گردید. به یاد بیاورید که هر نورون بیشتر برای آغاز حرکات چشم با جهت خاصی فعال است (برای مثال 10 درجه به سمت راست). زمانیکه این جهت بطور آزمایشی شناسایی شد، یک هدف انتخابی در آن موقعیت (داخل ناحیه پاسخ) ارائه شد و دیگر هدف انتخابی در موقعیت وارونه (مخالف ناحیه پاسخ یا

10 درجه به سمت چپ در این نمونه) ارائه شد. دوم، یک دوره هشدار زمانی بین حذف نقطه ثابت و ارائه اهداف انتخابی اعمال شد. بنابراین، میمونها در طول آزمایش یاد گرفتند که کجا و چه زمانی اهداف ارائه می‌شوند. این امر با الزاماتی همراه است که انتخاب حرکتی سریعاً بعد از ارائه تکمیل شود و انتخاب گزینه در طول دوره هشدار پیش برود. انتخابهای رفتاری به نسبت مساوی به هر هدف در یک الگوی غیرقابل پیش‌بینی اختصاص یافت، که نشان دهنده الگوهای رفتاری است که لی و همکاران (2004، شکل 26.4) مشاهده کردند. بررسی فعالیت عصبی SC نشان داد که حرکت چشم بطور فزاینده‌ای در زمانی که زمان ارائه هدف نزدیک می‌شود انتخاب می‌گردد (شکل 26.8). جالب توجه است که این فرآیند انتخاب عصبی، در فرآیند دیده شده در تصمیم‌گیری ادراکی زمانیکه فعالیت عصبی بعنوان تابعی از کیفیت شواهد حسی افزایش می‌یابد، دیده می‌شود. این نشان می‌دهد که اصول مشابهی که پایه مدل‌های انباشتگر مشخص هستند (فصل 3 و 19) برای هر دو شکل ادراکی و اجتماعی تصمیم‌گیری اعمال می‌شوند. به عبارت دیگر، درجه‌ای که فعالسازی‌های عصبی در طول زمان از بین می‌روند، بینشی را در دوره زمان انتخاب پاسخ قبل از اقدامات استراتژیک فراهم می‌آورد. در واقع، اگر طول مدت هشدار تغییر پیدا کرد، نرخ انتخاب‌پذیری عصبی نیز مطابق آن تغییر می‌کند.

برای درک این که چگونه این فرآیند انتخاب نورونی به نفع یکی از اهداف انتخابی در سطح SC تمایل پیدا می‌کند، نیازمند اندازه‌گیری فعالیت عصبی نواحی مدار حرکتی هستیم تا ورودیها به SC را فراهم کنیم. ابونفسا و دوریس (2011)، فعالیت نواحی چشمی قدامی (FEF) را ثبت کردند، در حالیکه میمونها فعالیت تطبیق پول را انجام می‌دادند. FEF قویا در ارتباط با بخشهای بزرگ لوبهای پیشانی و پارتيال است و ورودیهای قوی را به داخل SC فراهم می‌کند. علاوه بر این، FEF بطور خاصی در طول حرکات داوالبانه فعال است، بنابراین آنها را کاندیدای احتمالی درگیر در انتخاب حرکات چشم در طول فعالیت تطبیق پول با استراتژی مخلوط قرار می‌دهد. ابونفسا و دوریس در همان فعالیت و میمونها مطالعات SC را انجام دادند، بنابراین هر تفاوتی در پردازش عصبی بین FEF و SC بعید است که در نتیجه تفاوت در استراتژیهای رفتاری باشد. این نویسندگان دریافتند که انتخاب‌پذیری عصبی زودتر رخ داد و به سطح کلی بالاتری در طول زمان منجر به ارائه اهداف انتخابی در FEF نسبت به SC رسید (شکل 26.8). ممکن است انتظار

داشته باشید که انتخاب‌پذیری عصبی در SC بیشتر باشد زیرا نزدیکتر به خروجی حرکتی نهایی است و اطلاعات را در سراسر نواحی پیشانی آهیانه‌ای که در بالا توصیف شد ادغام می‌کند. یک توضیح احتمالی برای این امر این است که انتخاب‌پذیری عصبی در FEF فرآیند تصمیم‌گیری جاری را منعکس می‌کند، زیرا سطح آستانه‌ای که فعالیت عصبی باید برای آغاز حرکت داشته باشد در SC واقع است، این اطلاعات تصمیم یا به SC منتقل نمی‌شود یا SC بطور جزئی برای ارائه اهداف انتخابی در جهت جلوگیری از عبور زودگذر حرکات آستانه و زودرس مهار شده است.



شکل 26.8. انتخاب‌پذیری عصبی SC و FEF حرکات با استراتژی مخلوط آتی در طول فعالیت تطبیق پول. (آ) تصاویر خطوط موازی (پنلهای بالا) و توابع فعالسازی پس سیناپسی (پنل پایین) بر اساس حرکات مستقیم (سیاه) و مخالف (خاکستری) ناحیه پاسخ نورون. (ب) تکامل پیش‌بینی عصبی در طول زمان. تحلیل ویژگیهای راه‌اندازی گیرنده برای SC (نورون‌های N=78؛ آبی) و FEF (نورون‌های N=28؛ قرمز). حلقه‌ها پیش‌بینی‌های عصبی را بر اساس بسته‌های زمانی 50 میلی ثانیه موفق در طول دوره هشدار ارائه می‌دهند. پاسخهای عصبی جمعیت (C-D) در

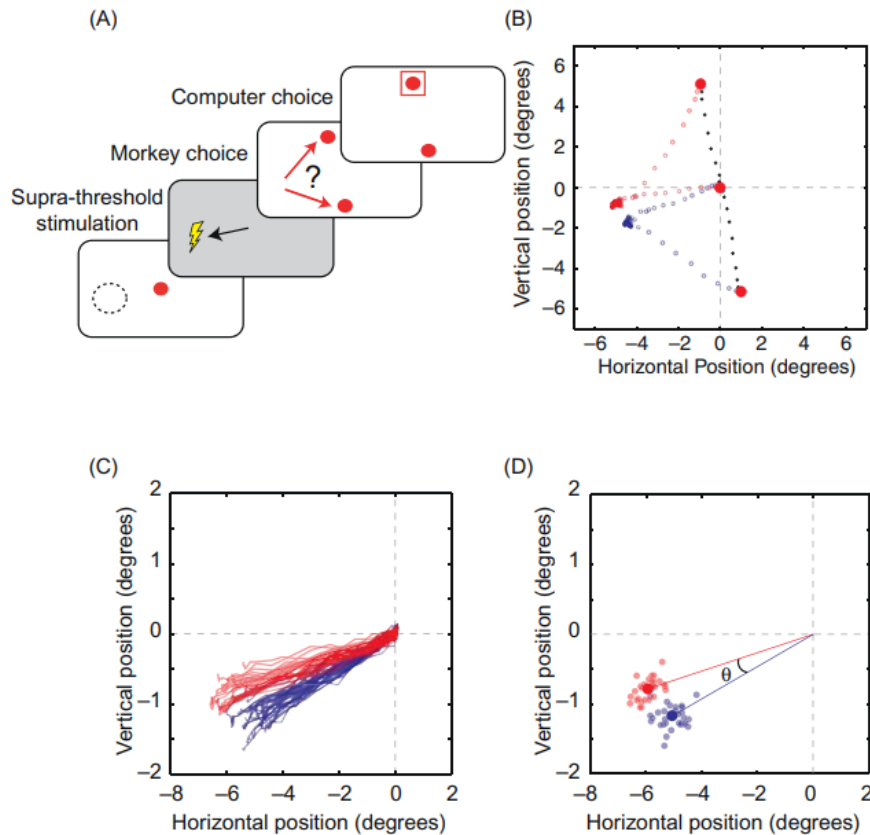


طول فعالیت تطبیق پول برای SC (ج) و FEF (د). خطوط سیاه فعالیت عصبی را زمانیکه هدف در ناحیه پاسخ نوروون انتخاب گردید را نشان می‌دهند و خطوط روشن نیز این فعالیت را در زمانی که هدف مخالف نواحی پاسخ نوروون انتخاب شد را ارائه می‌دهند.

این فعالیت پیش هدف در SC توسط تاریخچه انتخابهای قبلی و نتایج آنها در روندی مشابه با هرآنچه که در ساختارهای قشری بالاتر مشاهده شد مدوله می‌شود. یک گرایش پیروز ماندن آشکار است، اگر میمون حرکتی را انتخاب کند و در طول فعالیت تطبیق پول پاداش دریافت کند. تجمع سریعتر فعالیت در مکان خاصی بر روی نقشه SC با احتمال بیشتری برای انتخاب آن اقدام و زمانهای واکنش سریعتر همراه است. تفاراجاه و همکاران (2010) دریافتند که تخمین آزمون به آزمون ارزش اقدام حاصل از اعمال مدل EWA ترکیبی با فعالیت SC پیش هدف آزمون به آزمون همبستگی دارد. این مطالعه شواهد محکمی را فراهم آورد که رقابت بین جمعیت‌های عصبی در ساختارهای پیش حرکتی مغز در روندی پیش‌بینی شده توسط مدل‌های یادگیری برای انتخاب اقدامات استراتژیک تنظیم می‌گردد. اختلال مستقیم مدارهای عصبی نیز در فعالیت‌های تصمیم‌گیری برای فراهم آوردن شواهد کاربردی مرتبط با مشارکت ناحیه مغز برای رفتار انتخابی استفاده شده است.

با استفاده از پارادایم تحریک میکرو حاصل از مطالعه گولد و شادلن (2000، 2003؛ که در فصل 19 بحث شد)، تفاراجاه و همکاران (2009) این آزمایش را انجام دادند که آیا فعالیت پیش‌بینی در SC که در بالا نشان داده شد بصورت کارکردی با فرآیند انتخاب پاسخ تحت شرایط نظری بازی است (شکل 26.9). برای آزمایش این فرضیه، بر روی بخش کوچکی از آزمون‌های بازی تطبیق پول، فرآیند تصمیم‌گیری جاری با شروع میکرو تحریک دچار اختلال شد (شکل 26.9). این موقعیت SC تحریک شده حرکات متعادل به سمت اهداف انتخابی را منجر گردید. از آنجا که مسیرهای توسط فعالیت جمعیت در سراسر نقشه SC تعیین می‌شوند، حرکات تحرکی-القایی به سمت نواحی پیش از فعالیت منحرف می‌شوند، بطور موثری نشان می‌دهد که میمون چه گزینه‌ای را در فرآیند انتخاب برمی‌گزیند. نویسندگان دریافتند که این حرکات تحرکی-القایی به سمت موقعیتی که حیوان در نهایت انتخاب می‌کند منحرف شدند (شکل 26.9). زمانیکه تحریک نزدیکتر به زمانی که اهداف انتخابی ارائه شدند اعمال گردید، انحرافات بیشتر

واضح شدند. الگوی انحرافات تحریکی-القایی در طول زمان، دوره زمانی فرآیند انتخاب عصبی را از SC این میمونها ردیابی کرد (شکل 26.8). از اینرو، اختلال در توسعه برنامه‌های حرکتی چشم در محدوده زمانی پیش از ارائه اهداف انتخابی پنجره جدیدی را به دوره زمانی فرآیند انتخاب پاس تدریجی در طول تصمیم‌گیری با استراتژی مخلوط باز کرد (شکل 26.9).



شکل 26.9. استفاده از تحریک بیش از آستانه برای آزمودن نقش SC در آماده‌سازی حرکات چشم با استراتژی مخلوط. (آ) فعالیت رفتاری. تحریک اعمال شده در طول دوره هشدار باعث شروع حرکتی شد که قائم (متعامد) به اهداف بود. پس از آن، میمونها آزاد بودند که هر یک از دو هدف را انتخاب کنند. حلقه خالی و جرقه بردار حرکت تحریکی را نشان می‌دهند. حرکات (B-D) بر روی آزمونهای تحریک بر اساس انتخاب هدف نهایی جدا شدند. داده‌ها برای آزمونهایی نشان داده می‌شوند که در آنها تحریک 500 میلی‌ثانیه به داخل 600 میلی‌ثانیه دوره هشدار برای مکان تحریک مورد نظر اعمال شد. (ب) در اکثر آزمونها، تحریک اعمال نشد و حرکات چشم مستقیماً به سمت اهداف دوخته شد (سیاه). حرکات تحریکی به دو گروه تقسیم شدند: آنهایی که در آنها هدف سمت چپ (قرمز) و

سمت راست (آبی) در نهایت انتخاب شدند. (ج) حرکات تحریکی تمایل داشتند که به سمت هدفی که در نهایت انتخاب شد منحرف شوند. (د) انحراف زاویه‌ای بصورت زاویه بین ماینگین نقاط پایانی بین دو گروه از حرکات تحریک شده محاسبه شد.

در نهایت، تفاراجاه و همکاران (2009) تحریک زیر آستانه را برای SC در زمان منجر به انتخابهای حرکتی در فعالیت تطبیق پول به کار بردند. این تحریک سطح پایین برای فعالیت گرایشی در مکان تحریک SC کافی بود، اما برای شروع مستقیم حرکات کفایت نمی‌کرد. نتیجه این بود که استراتژی میمونها از تعادل ناش پیش‌بینی شده پاسخهای یکسان به دو هدف به نفع پاسخهای به سمت مکان تحریک تغییر پیدا کرد. این امر شواهدی را نشان داد که SC در انتخاب حرکات استراتژی مخلوط دخالت دارد و بیشتر این مورد را برجسته می‌کند که چگونه فعالیت مصنوعی در مدارهای انتخاب می‌تواند بینشی را در نقش کاربردی که ناحیه مغزی خاص در فرآیند تصمیم‌گیری بازی می‌کند را فراهم آورد.

### نتیجه‌گیری

این فصل پیشرفتهای مهمی را که در فهم مدارهای عصبی تصمیم‌گیری اجتماعی دخیل هستند را با ترکیب تکنیکهای نوروفیزیولوژیکی در راسته پستانداران غیر انسان با فعالیتهای میکرواقتصادی و تحلیلهای آماری صورت گرفته است را نشان داده است. تکنیکهای تهاجمی استفاده شده در پستانداران غیر انسان به فعالیت عصبی این اجازه را می‌دهد که با تفکیک پذیری فضایی و زمانی بالایی و بصورت همبسته با مراحل خاصی از بازی، رفتار و پارامترهای مدل‌های یادگیری ثبت شود. علاوه بر این، قابلیت الگوهای محلی فعالیتهای عصبی در بازی می‌تواند از طریق دستکاری مصنوعی بررسی شود. این تکنیکها به محققان این اجازه را داده است که ساختارهای پیشانی، عقده‌های قاعده‌ای و ناحیه بصری مغز را که برای تصمیم‌گیری اجتماعی ضرورت دارند، شناسایی کنند. این امر با تحلیل سیگنالهای رفتاری و عصبی در چارچوب مدل‌های یادگیری محقق شده است. این به ما اجازه می‌دهد که مکانیسم‌ها را با ارزش موارد ارائه شده طبق انتخابهای قبلی حیوان و نتایج آنها و نیز نحوه انتخاب آنها در اساس آزمون به آزمون درک کنیم.

هنگامی که آمارهای مهم انتخابها و نتایج در طول بازی خاصی محاسبه می‌شوند و مقادیر مختلفی از مدل‌های یادگیری در قشر پیشانی به روز می‌شوند، انتخاب واقعی و اجرای انتخاب باید صورت بگیرد. شواهد نشان می‌دهد که رقابتی بین جمعیت‌های عصبی در نواحی پیش حرکتی مغز (FEF و SC برای حرکات چشم) وجود دارد که اقدامات در اختیار را ارائه می‌دهد. بطور معمول، فعالیت در یک جمعیت برای غلبه بر دیگر جمعیتها آغاز می‌شود، زمانی که سطح آستانه فعالسازی فرا می‌رسد، یک انتخاب یا حرکت صورت می‌گیرد. به نظر می‌رسد که رقابت مشابهی برای تصمیمات صرفاً ادراکی در جایگاه آستانه عمل تحت تاثیر کیفیت اطلاعات حسی قرار می‌گیرد، رخ می‌دهد. برای تصمیمات اجتماعی، رقابت توسط عوامل اقتصادی مانند ارزش هدف، تاریخچه انتخابهای گذشته و نتایج آنها و حتی اطلاعات ساختگی در مورد اینکه چه نتایجی می‌تواند با خود داشته باشد، شکل گرفته است. اگرچه کار بیشتری باید انجام شود، آمارهای با درجه بالاتر و پارامترهای یادگیری رمزگذاری شده در شبکه‌های پیشانی آهیانه‌ای برای شکل دادن به رقابت در نقشه‌های سازمان یافته فضایی اقدامات بالقوه مانند آنهایی که در FEF و SC به نفع انتخاب با بیشترین ارزش برای انتخابگر گرایش دارند، ظاهر می‌شود. این حقیقت که مدارهای عصبی ذاتا شلوغ هستند، ممکن است برای تصمیم‌گیری اجتماعی مفید واقع شود؛ می‌تواند منبعی از بخت و اقبال باشد که این اطمینان را بدهد که با ارزشترین اقدام احتمال بیشتری دارد (اما قطعی نیست) که رخ بدهد. بنابراین، ممکن است مدارهای مغز ما برای تصمیم‌گیری اجتماعی در جهت بهره‌برداری از گزینه‌های با بیشترین ارزش در طول بازی و در زمان تزریق برخی از تصادفات برای جلوگیری رقیب از بهره‌مندی ما طراحی شده باشند.

## References

- Abe, H., Lee, D., 2011. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron*. 70, 731741.
- Abunafessa, A., Dorris, M.C., 2011. Role of the frontal eye fields in choosing mixed-strategy saccades. *Soc. Neurosci. Abstr.*
- Andersen, R.A., 1995. Encoding of intention and spatial location in the posterior parietal cortex. *Cereb. Cortex*. 5, 457469.
- Barraclough, D.J., Conroy, M.L., Lee, D., 2004. Prefrontal cortex and decision-making in a mixed-strategy game. *Nat. Neurosci.* 7, 404410.
- Behrens, T.E., Woolrich, M.W., Walton, M.E., Rushworth, M.F., 2007. Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 12141221.
- Belova, M.A., Paton, J.J., Salzman, C.D., 2008. Moment-to-moment tracking of state value in the amygdala. *J. Neurosci.* 28, 1002310030.
- Bernacchia, A., Seo, H., Lee, D., Wang, X.J., 2011. A reservoir of time constants for memory traces in cortical neurons. *Nat. Neurosci.* 14, 366372.
- Bisley, J.W., Goldberg, M.E., 2003. Neuronal activity in the lateral intraparietal area and spatial attention. *Science*. 299, 8186.
- Byrne, R., Whiten, A., 1989. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford Univ Press.
- Cai, X., Kim, S., Lee, D., 2011. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron*. 69, 170182.
- Camerer, C., Ho, T.H., 1999. Experience-weighted attraction learning in normal form games. *Econometrica*. 67, 827874.
- Camerer, C.F., 2003. *Behavioral Game Theory*. Princeton Univ Press, Princeton.
- Carello, C.D., Krauzlis, R.J., 2004. Manipulation intent: evidence for a causal role of the superior colliculus in target selection. *Neuron*. 43, 575583.
- Corrado, G.S., Sugrue, L.P., Seung, H.S., Newsome, W.T., 2005. Linear-nonlinear-poisson models of primate choice dynamics. *J. Exp. Anal. Behav.* 84, 581617.
- Cui, H., Andersen, R.A., 2007. Posterior parietal cortex encodes autonomously selected motor plans. *Neuron*. 56, 552559.
- Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. Model-based influences on humans' choices and striatal prediction errors. *Neuron*. 69, 12041215.
- Dorris, M.C., Glimcher, P.W., 2004. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*. 44, 365378.
- Dorris, M.C., Olivier, E., Munoz, D.P., 2007. Competitive integration of visual and preparatory signals in the superior colliculus during saccadic programming. *J. Neurosci.* 27, 50535062.
- Dorris, M.C., Pare, M., Munoz, D.P., 1997. Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. *J. Neurosci.* 17, 85668579.
- Erev, I., Roth, A.E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* 88, 848881.
- Fudenberg, D., Levine, D.K., 1998. *The Theory of Learning in Games*. MIT Press, Cambridge.
- Glimcher, P.W., 2003. The neurobiology of visual-saccadic decisionmaking. *Annu. Rev. Neurosci.* 26, 133179.