بسم الله الرحمن الرحيم

دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر

مدل‌سازی محاسباتی سوء مصرف مواد

نگارش:
امیر دزفولی

استاد راهنما:
دکتر کارو لوکس

استاد مشاور:
دکتر آذرخش مکری

پایان‌نامه برای دریافت درجهٔ کارشناسی ارشد در رشتهٔ
مهندسی کامپیوتر، گرایش هوش ماشین و رباتیک

خرداد ۸۸

صفحه تصویب پایان نامه توسط داوران: پس از دفاع از دفتر تحصیلات تکمیلی دانشکده دریافت شده و به جای این صفحه قرار می‌گیرد.

دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

عنوان:

مدل‌سازی محاسباتی سوء مصرف مواد

نگارش:

امیر دزفولی

پایان‌نامه برای دریافت درجهٔ کارشناسی ارشد در رشتهٔ

مهندسی کامپیوتر، گرایش هوش ماشین و رباتیک

از این پایان‌نامه در تاریخ در مقابل هیئت داوران دفاع به عمل آمد و مورد تصویب قرار گرفت.

معاونت آموزشی و تحصیلات تکمیلی پردیس دانشکده‌های فنی	دکتر جواد فیض
رئیس دانشکده	دکتر پرویز جبه‌دار مارالانی
سرپرست تحصیلات تکمیلی دانشکده	دکتر سعید نادر اصفهانی
استاد راهنما	دکتر کارو لوکس
استاد مشاور	دکتر آذرخش مکری
عضو هیئت داوران	دکتر مجید نیلی احمدآبادی
عضو هیئت داوران	دکتر مسعود اسدپور
عضو هیئت داوران	دکتر محمدرضا زرین‌دست

تعهد نامه اصالت اثر

اینجانب ..... تائید مي كنم كه مطالب مندرج در این پایان نامه حاصل كار پژوهشي اینجانب است و به دستاوردهاي پژوهشي دیگران كه در این نوشته از آنها استفاده شده است مطابق مقررات ارجاع گردیده است. این پایان نامه قبلاً براي احراز هیچ مدرك هم سطح یا بالاتر ارائه نشده است.

كلیه حقوق مادي و معنوي این اثر متعلق به دانشكده فني دانشگاه تهران مي باشد.

نام و نام خانوادگي دانشجو :

امضاي دانشجو :

تقدیم به پدر و مادر عزیزم

تشکر و قدردانی

از خداوند متعال متشکرم که مرا یاری کرد پژوهش حاضر را انجام دهم. همچنین از پدر و مادرم که در طول دوره تحصیل، هیچ‌گونه حمایتی را از من دریغ نکردند، نهایت سپاس را دارم.

آشنایی من با مباحث علوم شناختی، در درس هوش‌مصنوعی گسترده‌ی دکتر مجید نیلی احمدآبادی بود؛ و نحوه‌ی ارائه درس موجب علاقه‌مندی من به مباحث علوم‌شناختی و عصب‌شناسی شد که مقدمه‌ای بود برای ورود من به کار تحقیقاتی در این زمینه؛ از ایشان تشکر و قدردانی می‌کنم.

همچنین از خانم دکتر لاله قدک‌پور که همواره در خلال صحبت‌ها و جلسات با ایشان، با دیدگاهی جدید نسبت به ادامه مسیر آشنا می‌شدم، سپاسگزارم. دکتر بابک نجار اعرابی، در طول درس‌هایی که با ایشان گذراندم، با ابزارهای ریاضی مورد نیازم آشنا شدم و از این بابت از ایشان تشکر می‌کنم.

دکتر کارو لوکس، استاد راهنمای من بودند که حقیقتا مطالب زیادی از ایشان آموختم؛ از ایشان سپاسگزار و متشکرم. از دکتر آذرخش مکری به خاطر راهنمایی‌هایشان ممنونم و همچنین از دکتر حامد اختیاری به خاطر همکاری دوستانه و ثمربخش، سپاسگزارم.

دکتر سرج آهمد، دکتر دیوید ردیش، دکتر بوریس گوتکین، دکتر یال نیو و دکتر ناتانایل داو، که در طول انجام پایان‌نامه بحث‌های بسیار مفیدی با آنها داشتم و صبر و دقت و توجه آنها در پاسخ به سوالات من حیرت آور بود؛ از آنها متشکرم.

دوستان خوبم، پیام پیرای و محمد مهدی کرامتی، که در طول انجام پایان‌نامه از همه جهت به من کمک کردند؛ از آنها ممنونم. همچنین از محمد افشار که زحمت مرور پایان‌نامه را تقبل کرد و محمد-مهدی اجل لوئیان و حبیب کرباسیان که مقالات انگلیسی را مرور کردند، سپاسگزارم.

چکیده

اعتیاد را می‌توان به عنوان رفتاری اجباری و کنترل نشده برای جستجو و مصرف ماده‌ی اعتیادآور بیان کرد. یک تئوری اعتیاد، بر پایه‌ی مداخلات مواد بر ساختارهای عصبی یا روانی، توضیحی برای رفتارهای معتادانه ارائه می‌کند. در مدل‌سازیِ محاسباتیِ اعتیاد، یک تئوری اعتیاد توسط زبان ریاضیات توصیف می‌شود. این نوع مدل‌سازی امکان ارائه‌ی پیش‌بینی‌های کمّی و توصیف منسجم داده‌ها را فراهم می‌کند. این‌گونه مدل‌سازی که به عنوان واسطی بین داده‌های ساختاری (عصبی یا روانی) و داده‌های رفتاری عمل می‌کند، در حوزه‌ی پیش‌بینیِ رفتار و درمان کاربرد دارد.

پژوهش حاضر شامل دو بخش است. در بخش اول، بر اساس یافته‌های عصبی و با استفاده از مدل یادگیری تقویتیِ متوسطِ پاداش، مدلی عصبی-محاسباتی برای اعتیاد ارائه شده است. مدل پیشنهاد-شده با داده‌های حاصل از مدل‌های حیوانیِ جستجو و مصرف اجباری مواد منطبق است. همچنین مدل پیشنهادشده، برخلاف مدل‌های پیشین، تکانشگری و کاهش انگیزه را برای پاداش‌های طبیعی پس از مصرف درازمدت مواد، توضیح می‌دهد. از نظر ساختاری مدل بیان‌شده، بر اساس سه فرض عصبی بنا شده است. اول آنکه میزان فعالیت فازیک نورون‌های دوپامینی، خطای پیش‌بینی را کد می-کند. دوم، مواد اعتیادآور به طور غیرطبیعی میزان دوپامین را زیاد می‌کنند. سوم، مصرف مزمن مواد اعتیادآور (به علت کم شدن گیرنده‌های دوپامینی و یا بالارفتن سطح فعالیت یکنواخت نورون‌های دوپامینی در استریاتوم) موجب بالا رفتن آستانه‌ی تحریک سیستم پردازش پاداش می‌شود.

در بخش دومِ این پژوهش، نظام تصمیم‌گیری افراد معتاد و گروه شاهد در شرایط مخاطره‌آمیز، با استفاده از آزمون قمار آیوا مورد ارزیابی قرار گرفته است. نتایج بدست آمده نشان می‌دهد که هر دو گروه معتادِ درمان‌جو (n=217) و شاهد (n=130) عملکرد ضعیفی در آزمون داشتند و امتیاز خالص آنها کمتر از ۱۰ بود. همچنین عملکرد گروه معتاد نسبت به گروه شاهد ضعیف‌تر بود ($p < 0.05$). به منظور کشف علت ضعف دو گروه، از مدل‌سازی شناختی استفاده شد. در این راستا، با برازش مدل‌های

مختلفی از تصمیم‌گیری بر داده‌های هر دو گروه به طور جداگانه، مدل بهنیه برای هر گروه مشخص شد. نتایج به‌دست آمده نشان می‌دهد در هر دو گروه، مدلی که برای ارزیابی انتخاب‌های مختلف صرف‌نظر از مقادیر پاداش و زیان تنها به تعداد آنها توجه می‌کند، مدل بهینه است. همچنین استفاده از روش تحلیل حساسیت مشخص کرد که مدل بهینه برای گروه معتاد، علاوه بر توجه صرف به تعداد دفعات پاداش و زیان، بیشتر از گروه شاهد تحت تاثیر آسیب‌گریزی قرار دارد. نتایج فوق گویای این مطلب است که علت کم بودن امتیاز کسب شده در هر دو گروه به خاطر نادیده گرفتن اندازه‌ی پاداش و زیان و تنها توجه به دفعات آن است. علاوه بر آن پایین بودن بازده در گروه معتاد نسبت به گروه شاهد به علت عدم تعادلِ آسیب‌گریزی، پاداش جویی و گرایش به سمت آسیب‌گریزی است.

فهرست مطالب

فهرست جدول‌ها

فهرست شکل‌ها

فهرست اختصارها

اختصار معادل	عبارت
NA	nucleus accumbens
VTA	ventral tegmental area
PFC	prefrontal cortex
SNc	substantia nigra pars compacta
fMRI	functional magnetic resonance imaging
PET	positron emission tomography
DDT	delayed discounting task
CNS	central nervous system
FR	fixed-ratio schedule
PR	progressive-ratio schedule
US	unconditioned stimulus
CR	conditioned response
nAChR	nicotine acetylcholine receptor

فصل ۱

اعتیاد – مدل‌سازیِ اعتیاد

۱-۱ مقدمه

در این فصل ابتدا به ارائه‌ی یک تعریف رفتاری برای اعتیاد می‌پردازیم. در مرحله‌ی بعد مسئله‌ی سوءِ-مصرف مواد را در سطح کلان بیان کرده و رویکردهای مبتنی بر مدل‌سازیِ محاسباتی[۱] را برای حل این مسئله مرور می‌کنیم. سپس به ارائه‌ی مسئله در سطح خرد- که موضوع پژوهش حاضر است- می‌پردازیم و رویکرد انتخاب شده در این پژوهش را به همراه روش‌شناسی آن شرح می‌دهیم.

۱-۲ ماده‌ی اعتیادآور، معتاد

اعتیاد یک بیماری مغزی است که بر اثر تاثیرگونه‌ی خاصی از مواد شیمیایی، که مواد اعتیادآور نامیده می‌شوند، بر نظام عصبی مرکزی[۲] ناشی می‌شود[۳]. مواد اعتیادآور با تحت تاثیر قرار دادن عملکرد زیستی در انسان و حیوان و با ایجاد اختلال در نظام عصبی مرکزی باعث بروز رفتار معتادانه[۴] می-

---

[1] Computational modeling

[2] Central Nervous System (CNS)

[3] در راهنمای تشخیصی و آماری انجمن روانپزشکی آمریکا شماره پنج، واژه وابستگی به مواد (drug dependence) به جای اعتیاد تصویب شده است [۸۹]. لیکن در این پایان‌نامه به علت مرسوم بودن آن، از واژه‌ی اعتیاد استفاده می‌شود.

[4] Addictive behavior

شوند. راهنمای تشخیصی و آماری انجمن روانپزشکی آمریکا (DSM-IV)، هفت معیار و علامت برای وابستگی به مواد ذکر می‌کند (شکل ۱-۱). این راهنما وجود حداقل سه نشانه از هفت نشانه را برای تشخیص وابستگی لازم می‌داند. در راهنمای بیان شده انواع مواد مخدر نیز بیان شده‌اند [۱] که این مواد عبارتند از الکل، کافئین[۱]، حشیش[۲]، توهم زاها[۳]، هروئین[۴] و دیگر افیونها[۵]، مواد استنشاقی[۶]، نیکوتین[۷]، فن سیکلیدن[۸]، مسکنها[۹]، محرک زاها[۱۰] (کوکائین[۱۱] و آمفتامین‌ها[۱۲]).

مکانیزم عمل مواد[۱۳] مختلف با یکدیگر متفاوت است و می‌توانند باعث ایجاد آثار مختلفی در فرد شوند. به عنوان مثال می‌توانند نشانه‌های ترک متفاوتی داشته باشند یا حتی در بین افراد مختلف عارضه‌های متفاوتی ایجاد کنند. با این وجود، نقطه‌ی مشترک همه‌ی آنها به لحاظ رفتاری ایجاد نشانه‌های بیان شده‌ی فوق است.

در صورتی که وجود نشانه‌های بیان شده را برای اعتیادی بودن گونه‌ای از مصرف کافی بدانیم، می‌توان اعتیاد را به برخی از فعالیت‌های شناختی‌تر نیز گسترش داد. به عنوان مثال یک معتادِ به قمار[۱۴] رفتاری مشابه با اعتیادِ به مواد از خود نشان می‌دهد. همچنین برخی از مختصصان گونه‌ای از رفتارهای جنسی و یا وابستگی به محصولات تصویری جنسی را نیز نوعی اعتیاد می‌دانند و در قالب اعتیاد جنسی[۱۵] و یا اعتیاد به تصاویر جنسی[۱۶] تلقی می‌کنند [۲]. به طور مشخص، اعتیاد به اینترنت در راهنمای تشخیصی و آماری انجمن روانپزشکی آمریکا، شماره پنج (DSM-V) تعریف شده، نشانه-های آن ذکر شده [۳] و در قالب سه عملِ بازی کامپیوتری بیش از حد، استفاده‌ی جنسی بیش از

[1] Caffeine
[2] Cannabis
[3] Hallucinogens
[4] Heroin
[5] Opiates
[6] Inhalants
[7] Nicotine
[8] Phencyclidine
[9] Sedative
[10] Stimulants
[11] Cocaine
[12] Amphetamine

[۱۳] در این پایان نامه از واژه مواد به معنای مواد اعتیادآور استفاده شده است.

[14] Gambling
[15] Sexual addiction
[16] Pornography addiction

ملاک‌های وابستگی به مواد (بیان شده در DSM-IV)

۱. تحمل؛
   - نیاز بیش از پیش به مواد برای رسیدن به تاثیر مطلوب،
   - کاهش تاثیر مواد به مرور زمان در مقایسه با میزان مشابه در قبل.
۲. ترک؛
   - بروز علائم ترک،
   - استفاده از ماده مصرفی و یا مشابه آن برای کاهش علائم ترک.
۳. مصرف بیشتر و طولانی‌تر از آنچه که شخص انتظار و قصد انجام آن را دارد.
۴. بی‌نتیجه بودن تلاش‌های فرد برای کاهش و یا ترک مواد.
۵. صرف مدت زمان زیاد برای به دست آوردن ماده.
۶. کنار گذاشتن فعالیت‌های مهم اجتماعی، کاری و یا تفریحی به علت مصرف مواد.
۷. مصرف مستمر مواد برخلاف علمِ معتاد به اثرات سوء جسمی، روانی و اجتماعیِ مصرف.

شکل ۱-۱ علامت‌های اعتیاد از دیدگاه DSM-IV که وجود حداقل سه نشانه از هفت نشانه به منظور تشخیص وابستگی به مواد ضروری است.

اندازه و فرستادنِ زیاد پیامک و نامه‌ی الکترونیکی بیان شده است. مشکل اساسی برای مطالعه‌ی این-گونه وابستگی‌ها، محدود بودن به انسان، وجود نداشتن مدل‌های حیوانی برای آنها، و در نتیجه تا حدودی ناشناخته ماندن زیرساخت‌های عصبی این گونه وابستگی‌ها است.

همانطور که از نشانه‌های بیان شده برای اعتیاد آشکار است، وابستگی به مصرف، مواد نظام تصمیم-گیریِ فرد را تحت تاثیر قرار می‌دهد. از آنجایی که بسیاری از فعالیت‌های اجتماعی وابسته به چگونگی تصمیم‌گیریِ افراد است، وابستگی به مواد از حالت یک بیماری فردی خارج شده و تبعات اجتماعی وسیع پیدا می‌کند. این امر هنگامی که سوءمصرف مواد شیوع پیدا کرده و گسترده باشد، صورتی خاص پیدا می‌کند که در ادامه به آن اشاره می‌کنیم.

۱-۳ اعتیاد در نگاه کلان

سابقه‌ی مصرف مواد اعتیادآور به دوران پیش از تاریخ بازمی‌گردد. از میلیون‌ها سال پیش بشر به خاصیت ویژه برخی گیاهان پی برده و از آنها به شکلی ویژه استفاده می‌کرده است. به عنوان مثال آنها خشخاش[1] را هدیه‌ای آسمانی از طرف خدایان می‌دانستند که موجب تسکین دردهایشان می‌شود [۴]. از اوایل قرن بیستم این مصرف شکل تازه‌ای به خود گرفته، به شدت گسترش یافته و به صورت یک اپیدمی درآمده است. این گسترش مصرف، آن را از صورت یک امر تفریحی و محلّی خارج کرده و به یک مسئله‌ی بین‌المللی تبدیل کرده است.

از دیدگاه اجتماعی، مواد اعتیادآور مشکلات سلامت متعددی را به همراه دارند. اینگونه مشکلات بر پایه‌ی ویژگی‌های رفتاری یک معتاد که در بخش قبل بیان شد، مورد انتظار است. از آنجایی که معتاد با وجود آگاهی از عواقب ناگوار مصرف، قادر (یا مایل) به ترک رفتار خود نیست (علامت ۷ و ۴)، ممکن است به برخی از رفتارهای خطرناک، از قبیل استفاده از سرنگ فرد دیگری اقدام کند که این امر باعث انتقال و گسترش بیماری‌هایی مانند ایدز می‌شود. آمارها نشان می‌دهد مصرف مواد اعتیادآور با ایجاد مشکلات سلامت، از قبیل انتقال ویروس ایدز، باعث مرگ حدودا ۵.۳ میلیون نفر در سال می‌شود. از طرفی الگوی خاص مصرف این مواد مانند رشد صعودی مصرف در یک فرد (علامت ۱) و مصرف‌کننده‌ی دائمی بودن فرد معتاد (علامت ۴) باعث سودآوری بسیار زیاد تجارت این مواد شده است. گردش مالی تجارت این مواد سالیانه بالغ بر ۸۵ بیلیون دلار است که این رقم بیشتر از تولید خالص ملی ۳۴ درصد از کشورهای جهان است [۵]. از دیدگاه جرم‌شناسی، این تجارت غیر قانونی باعث افزایش جرم در جامعه می‌شود. از دیدگاه کار، مصرف این مواد باعث کاهش کارایی، غیبت از کار و مشکلات دیگری از این قبیل می‌شود که در کل باعث کاهش کارایی نیروی کار جامعه می‌گردد (علامت ۶). مصرف مواد اعتیادآور موجب سلب آزادی فرد و مانع رشد مصرف کننده شده و به خصوص در میان نسل جوان باعث از بین رفتن این سرمایه مهم اجتماعی می‌شود.

[1] Cannabis

این‌گونه تبعات اجتماعیِ سوءِ مصرف مواد موجب شده این مشکل به یک معضل اجتماعی و دغدغه‌ی دائمی سیاست‌گذاران این حوزه تبدیل شود. به طور کلی در سطح کلان می‌توان مسئله‌ی سوء مصرف مواد را به صورت یافتن سیاست‌گذاری بهینه به منظور برآورده کردن حداکثریِ سه منظور زیر تعریف کرد:

۱. درمان افراد معتاد،
۲. جلوگیری از گرایش افراد جدید،
۳. کاهش آسیب‌های سوء مصرف مواد (در معتادان فعلی).

هر یک از این سه رویکرد به وسیله‌ی اعمال سیاست‌های خاصی در جامعه تحقق می‌یابد. به عنوان مثال به منظور درمان افراد معتاد می‌توان به توسعه‌ی مراکز درمانی توجه کرد. به منظور جلوگیری از گرایش می‌توان آموزش عمومی در این زمینه را تقویت کرد؛ و به منظور جلوگیری از آسیب می‌توان از سیاستِ توزیعِ سرنگ مجانی بهره جست. به مانند هر سیاست‌گذاری دیگری، در این حوزه نیز اعمال یک سیاست کارا نیازمند پیش‌بینی عواقب آن است. بدین معنا که سیاست‌گذار باید بداند که پس از اعمال یک سیاست، وضعیت مصرف به چه سمتی تغییر خواهد کرد. انجام چنین پیش‌بینی در حوزه‌ی سوء مصرف مواد اغلب ساده نبوده و با پیچیدگی‌های زیادی همراه است. به عنوان مثال در ایران انتظار می‌رفت که پس از جمع‌کردن خرده فروش‌های مواد، مصرف این مواد کاهش یابد. با این وجود پس از اعمال این سیاست نه تنها میزان مصرف مواد کم نشد، بلکه میزان جرم ناشی از تهیه مواد نیز زیادتر شد. این واقعه بدان علت بود که مصرف در معتادان همانطور که در نشانه‌ها ذکر شد، نسبت به هزینه‌ی تهیه و مصرف مواد حساس نیست. بنابراین گرچه با جمع کردن خرده فروش‌ها، تهیه‌ی مواد سخت‌تر شده بود، لیکن معتادان هزینه این سختی را پرداخت کرده و به روش‌های جایگزین مواد تهیه می‌کردند. از طرفی جمع کردن خرده فروش‌ها موجب افزایش قیمت مواد شده و در نتیجه باعث افزایش جرم برای تهیه هزینه مالی می‌شود[۱].

[۱] مستند منتشر نشده از علیرضا ساکت.

مثال فوق نشان می‌دهد به منظور شناخت و مطالعه‌ی اعتیاد توجه به چندوجهی بودن این پدیده ضروری است. در مرحله‌ی اول این پدیده از تاثیر شیمیایی یک ماده اعتیادآور بر بدن فرد مصرف کننده نشات می‌گیرد. در نتیجه‌ی این امر، در مرحله دوم رفتار فرد تغییر کرده و گونه‌ای خاصی از تصمیم‌گیری را از خود نشان می‌دهد. از آنجایی که فرد در یک بستر اجتماعی با افراد دیگر در ارتباط است (از قبیل مددکاران، پلیس، دلالان مواد و غیره) این پدیده از حالت فردی خارج شده و جنبه‌ی گروهی به خود می‌گیرد. با توجه به متعدد بودن عوامل دخیل به نظر می‌رسد که این پدیده به صورت خام و با در نظر گرفتن تمام عوامل قابل مطالعه نیست و نیاز به ساده‌سازی و تقسیم مسئله به سطوح مختلف است. برای این منظور می‌توان از رویکرد مدلسازی در سطوح مختلف بهره جست. این موضوع در بخش بعد بحث شده است.

۱-۴ مدل‌سازی اعتیاد

مدل‌سازی گونه‌ای از شناخت است که در آن برخی ویژگی‌های پدیده‌ی مورد بحث برجسته شده و از برخی دیگر صرف‌نظر می‌شود. در نهایت ویژگی‌های برجسته شده در قالب یک زبان مشخص بیان می‌شوند. زبان مدل‌سازی با توجه به پدیده‌ی مورد توصیف و میزان پیچیدگی آن می‌تواند کمّی یا کیفی باشد. در توصیف پدیده‌های پیچیده اغلب از گونه‌ای از مدل‌سازی به نام مدل‌سازیِ محاسباتی استفاده می‌شود. در این رویکرد از ریاضیات برای توصیف پدیده مورد بحث استفاده می‌شود و دهه‌ها است که در زمینه‌های مختلف علمی مانند زیست‌شناسی، اپیدمی‌شناسی، مطالعات جمعیت و اقتصاد مورد بهره‌برداری قرار گرفته است. به عنوان مثال در اپیدمی‌شناسی، مدل‌سازی محاسباتی این امکان را فراهم می‌کند که نتایج حاصل از یک سیاست‌گذاری و یا سناریوهای مختلف پیش‌گیری، شبیه‌سازی شود. به عنوان یک نمونه، مدل‌سازیِ ریاضیِ استراتژی‌های مختلف به دولت انگلستان کمک کرد که به نحو موثری با بیماری‌های پا و دهان مقابله کند [۶].

توانایی‌های این روش موجب شده است که این رویکرد در برخورد با پدیده اعتیاد نیز به کار گرفته شود و پروژه‌های مختلفی تاکنون در سراسر دنیا در این زمینه انجام شده است. با توجه به جنبه‌های فردی و اجتماعی این پدیده، می‌توان مدل‌سازی در این حوزه را به بخش‌های کلان، خرد-کلان و خرد

تقسیم‌بندی کرد (شکل ۱-۲؛ برای مرور مراجعه شود به [۷]). در ادامه به تشریح هر سطح مدل‌سازی می‌پردازیم.

۱-۴-۱ مدل‌های خرد

موضوع مورد مطالعه در سطح خرد، چگونگی رفتار فرد معتاد در شرایط مختلف (مدل‌سازی رفتاری) و سازوکار تاثیر مواد اعتیادآور بر نظام عصبی آن فرد (مدل‌سازی عصبی-محاسباتی و دارویی) یا چگونگی تاثیر مواد بر توانایی‌های شناختی فرد (مدل‌سازی شناختی) است. حوزه‌ی پژوهش حاضر مدل‌سازی خرد است که این نوع مدل‌سازی را در بخش ۱-۵ شرح می‌دهیم.

شکل ۱-۲ مدل‌های خرد، کلان و خرد-کلان در مدلسازی اعتیاد

۱-۴-۲ مدل‌های خرد-کلان

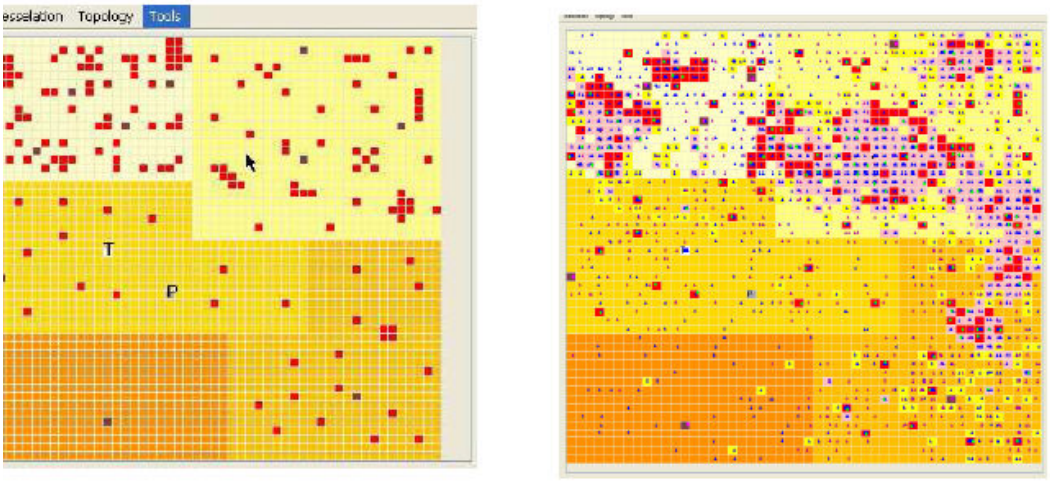
مدل‌های خرد-کلان بر خلاف مدل‌های خرد، در پی کشف رفتارِ جمعی و توصیف شاخص‌های جمعی عامل‌های موثر در اعتیاد هستند (مانند میزان مصرف کل مواد، میزان کل ارتکاب جرم و غیره.). روش مدل‌سازی به این گونه است که پس از شناسایی رفتار تک‌تک عامل‌ها، رفتار جمعی آنها در کنار یکدیگر شبیه‌سازی شده، از نتیجه‌ی آن شاخص‌های جمعی استخراج می‌شود. از پروژه‌های انجام شده در این زمینه می‌توان به پروژه موسوم به سیم‌دراگ[۱] اشاره کرد که در آن دوران قحطی هرویین در ملبورن مدل‌سازی شده است [۸]. این پروژه با شبیه‌سازیِ شهر ملبورن و عوامل اجتماعیِ درگیر (مانند معتاد، پلیس، درمانگر و غیره) توانسته که نقاط بحرانی (نقاطی که در آنها احتمال ارتکاب جرم زیاد است) و تعداد بیش مصرفی‌ها[۲] را پیش‌بینی کند (شکل ۱-۳).

کارکرد اصلی اینگونه مدل‌سازی و مدل‌سازی کلان در سیاست‌گذاری بهینه است؛ که نیاز به آن پیش از این بیان شد. مدل‌سازی کلان و خرد-کلان این امکان را فراهم می‌کنند که:

- در حوزه‌هایی که محدودیت داده وجود دارد، تخمینی قابل‌قبولی از شاخص مطلوب ارائه شود.
- نتایج حاصل از سیاست‌گذاری‌های مختلف پیش‌بینی و شبیه‌سازی شود.
- امکان سیاست‌گذاری پویا فراهم شود.

سه کارکرد فوق در واقع فراهم‌کننده‌ی یک ابزار پشتیبانی تصمیم‌گیری[۳] برای سیاست‌گذاران درحوزه‌ی اعتیاد است. با استفاده از مدل‌های توسعه‌یافته در حوزه‌ی کلان می‌توان روابط پیچیده‌ی میان اجزای درگیر در پدیده‌ی اعتیاد را به طور منسجم توصیف کرد و نتایج حاصل از هر سیاست-گذاری (مانند وضع قوانین حقوقی، قوانین مالی و غیره) را مورد بررسی قرار داد. نیاز به سیاست‌گذاریِ پویا و تامین آن توسط مدل‌سازی محاسباتی از این واقعیت نشات می‌گیرد که مصرف مواد در طی زمان در بین جمعیت‌های مختلف تغییر می‌کند و به همین نسبت یک سیاستِ کارا باید با این تغییرات متناسب باشد، که این همان نیاز به پویایی در حوزه سیاست‌گذاری مواد است. به عنوان مثال متوقف کردن منابع تامین مواد می‌تواند در مراحل اولیه‌ی یک اپیدمی، کارا باشد. این در حالی است

[۱] SimDrug
[۲] Overdose
[3] Decision support system



شکل ۱-۳ برگرفته از [۸]. نمونه‌ای از مدلسازیِ خرد-کلان اعتیاد در پروژه سیم‌دراگ. شکل سمت چپ شهر ملبوران را در ابتدای شبیه‌سازی نشان می‌دهد. نقاط قرمز نقاطی از شهر هستند که در آنجا احتمال انجام جرم بالا بوده است. شکل سمت راست شهر مبلورن را پس از شبیه‌سازی نشان می‌دهد. نقاط قرمز نشان‌دهنده‌ی مکان‌هایی از شهر است که در آنها احتمال وقوع جرم بالا پیش‌بینی می‌شود.

که در مراحل بعدی و پایانی اپیدمی، گرایش به سمت درمان کاراتر است. مدل‌سازی محاسباتی این امکان را فراهم می‌کند که روند تغییر اپیدمی در طول زمان پیش‌بینی شده و نتایج هر سیاست شبیه‌سازی شود.

۱-۴-۳ مدل‌های کلان

مدل‌های کلان همانند مدل‌های خرد-کلان در پی توصیف شاخص‌های اجتماعی هستند. تفاوتی که این گونه مدل‌سازی با دو نوع مدل‌سازی قبلی دارد در این است که این مدل‌سازی در سطح افراد انجام نمی‌شود و به طور مستقیم روابط حاکم بر شاخص‌های اجتماعی مطالعه می‌شوند. مدل‌های اقتصادی، مردم‌شناسی[۱] و اپیدمی‌پویا[۲] سه جریان حاکم بر این حوزه هستند. مدل‌های اقتصادی [۹,۱۰] با استفاده از مفاهیم عرضه و تقاضا و استفاده از قیمت مواد سعی در پیش‌بینی شاخص‌های بازار مانند تعداد مصرف‌کنندگان، قیمت و غیره دارند. رویکرد اپیدمیِ پویا [۱۱,۱۲] با استفاده از مدل‌های پویا، فرایند گسترش و شیوع یک ماده را در یک جمعیت شبیه‌سازی می‌کند. از پروژه‌های

[1] Ethnographic models
[2] Dynamical epidemic models

انجام‌شده در این زمینه می‌توان به پروژه‌ی مدل‌سازی تقاضای کوکائین در ایالات متحده اشاره کرد که مدل حاصل از آن قابلیت پیش‌بینی تعداد معتادین (به تفکیک میزان شدت اعتیاد) را دارد [۱۳].

۱-۵ مدل‌سازیِ خرد

همانطور که اشاره شد در حوزه‌ی مدل‌سازی خرد سه رویکرد مدل‌سازیِ رفتاری، مدل‌سازی شناختی و مدل‌سازی عصبی-محاسباتی (و دارویی) وجود دارد. در این پژوهش بر روی مدل‌سازی عصبی-محاسباتی و شناختی تمرکز شده که اولی موضوع بخش اول پایان‌نامه (فصل ۱ تا ۴) و دومی بخش دوم (فصل ۵) است. علت انتخاب این دو حوزه در بخش‌های بعدی و در حین بیان کارکردهای این نوع مدل‌سازی بیان می‌شود. در ادامه در ابتدا به توضیح مدل‌سازی عصبی-محاسباتی پرداخته و سپس مدل‌سازی شناختی را به طور اجمالی توضیح می‌دهیم.

۱-۵-۱ مدل‌سازی عصبی-محاسباتی اعتیاد

مدل‌سازی عصبی-محاسباتی[1] گونه‌ای از مدل‌سازی محاسباتی است که در دو سطح نورونی و سیستمی قابل انجام است. در سطح نورونی، در مدل مولفه‌هایی متناظر با فعالیت یک تک نورون وجود دارد؛ در حالی که در سطح سیستمی، متغیرها و مولفه‌های مدل به فعالیت مجموعه‌ای از نورون‌ها نسبت داده می‌شوند. مدل‌سازی انجام شده در این پژوهش در سطح سیستمی است.

اصول کلی روش مدل‌سازی عصبی-محاسباتی بدین صورت است که در ابتدا مجموعه‌ای از روابط ریاضی طراحی می‌شوند و ادعا می‌شود که این مجموعه از روابط مدلی عصبی-محاسباتی برای فعالیت بخشی از نظام عصبی ارائه می‌کنند. این ادعا به لحاظ علمی قابل قبول است اگر در درجه‌ی اول بتوان بین متغیرهای بیان شده در مدل و فعالیت بخش‌های مختلف مغز نگاشت برقرار کرد (صحت‌سنجی ساختاری) و در درجه‌ی دوم، اگر به نظام عصبی و مدل ورودی‌های یکسان داده شود، خروجی‌های مشابه حاصل شوند (صحت‌سنجی رفتاری).

در حوزه‌ی مدل‌سازی عصبی-محاسباتی اعتیاد، فرایند مدل‌سازی کمی پیچیده‌تر است. در قدم اول باید یک مدل از یک نظام عصبی سالم (نظام عصبی بدون تاثیر ماده اعتیادآور بر آن) ارائه شود. در

[1] Neurocomputational modeling

مرحله دوم باید نگاشتی بین متغیرهای مدل و زیر ساخت عصبی آن برقرار شود. در مرحله‌ی بعد باید شواهدی از چگونگی تاثیر ماده بر نظام عصبی ارائه شود. پس از این در صورتی که ماده بر ناحیه‌ای از نظام عصبی تاثیر می‌گذاشت که در دامنه‌ی نگاشت یکی از متغیرهای مدل قرار می‌گرفت، تغییر ایجاد شده در نظام عصبی به صورت تغییر مقدار متغیر متناظر مدل می‌شود. در مرحله‌ی آخر باید مدل حاصل شبیه‌سازی شده و نشان داده شود که با شواهد رفتاری اعتیاد سازگار است.

با توجه به فرآیند بالا، برای مقایسه‌ی مدل‌های مختلف نیاز است که شواهد رفتاری و عصبی مشخصی به عنوان معیار در نظر گرفته شود. خوشبختانه در حوزه‌ی اعتیاد وجود مدل‌های حیوانی باعث غنای داده در این حوزه هم در سطح رفتاری و هم در سطح عصبی شده است. به علاوه آزمایش‌های تصویربرداری مغزی متعددی نیز انجام شده که منبع مهمی برای داده‌های عصبی در انسان به شمار می‌روند. به منظور مقایسه‌ی مدل‌های مختلف در ابتدا لازم است که مجموعه‌ای از شواهد رفتاری و عصبی در نظر گرفته شوند و هر مدل بر حسب توانایی توصیف و پیش‌بینی مورد قضاوت قرار گیرد.

به لحاظ رفتاری می‌توان دو مولفه‌ی اصلی برای اعتیاد در نظر گرفت [۱۴]:

۱. رفتار اجباری مصرف و جستجوی مواد[1]؛

این رفتار بدان معناست که فرد معتاد با وجود عواقب سوء رفتار خود بر آن اصرار می‌ورزد. این ویژگی از مهمترین ویژگی‌های اعتیاد به شمار رفته و در DSM-IV به عنوان شاخصه اصلی اعتیاد معرفی شده است. به طور دقیق‌تر، در معتادان *پس از مصرف طولانی مدت*، رفتار مصرف و جستجوی مواد، حساسیت خود را نسبت به هزینه‌ی این کار از دست می‌دهد. یا به عبارت دیگر همزمان با پیشرفت مراحل اعتیاد، چسبندگی[2] مصرف مواد به هزینه‌ی این کار کم می‌شود. شواهد این امر در مدل‌های حیوانی در بخش ۴-۴-۳ توضیح داده شده است.

۲. اختلال در نظام پردازش پاداش‌های طبیعی؛

[1] Compulsive drug seeking and taking
[2] Elasticity

این اختلال بدان معناست که انگیزه‌ی معتاد برای کسب پاداش‌های طبیعی کاهش پیدا می‌کند. همچنین در تصمیم‌گیری‌های شامل پاداش‌های طبیعی، دچاره عارضه‌هایی مانند تکانشگری[1] می‌شود [۱۵].

یک تئوری کارای اعتیاد باید در سطح رفتاری با دو نشانه‌ی بالا منطبق باشد. معنای دقیق این دو نشانه در سطح محاسباتی در بخش‌های بعدی بررسی شده است.

در سطح عصبی، از آنجا که مکانیزم عصبی و دارویی مواد مختلف با یکدیگر متفاوت است، در مورد همه‌ی مواد نمی‌توان مجموعه‌ی یکتایی از شواهد عصبی را به عنوان معیار در نظر گرفت. در فصل‌های آینده با معرفی هر مدل به داده‌های عصبی زیربنای مدل پرداخته و شواهد عصبی استفاده‌شده در مدل بیان می‌شوند.

مدل‌سازیِ عصبی-محاسباتی اعتیاد مانند هر دو رویکرد دیگر (رفتاری و شناختی) با پیش‌بینیِ رفتار یک فرد می‌تواند تامین‌کننده‌ی مدل رفتاری یک عامل برای استفاده در شبیه‌سازی‌های خرد-کلان باشد. در سطح بالینی، مدل‌سازیِ عصبی-محاسباتی دارای دو کارکرد مهم است. کارکرد اول اینکه با توصیف یکپارچه‌ی داده‌های ساختاری (عصبی) و رفتاری، ارتباط میان این دو سطح را برقرار می‌کند. به عبارتی توضیحی در سطح عصبی برای رفتارهای مشاهده‌شده ارائه می‌دهد. ارائه‌ی این توضیح می‌تواند به درمان رفتار معتادانه و سیاست‌گذاری عمومی کمک کرده و در مواردی، تبعاتِ رفتاریِ یک دخالت ساختاری را در نظام عصبی پیش‌بینی کند. از طرفی محاسباتی بودن این رویکرد، امکان ارائه‌ی پیش‌بینی‌های کمّی را فراهم می‌کند؛ که این امر به سطح‌سنجی مدل کمک کرده و با ارائه پیش‌بینی‌های جدید مسیری برای آزمایش‌های عصبی آینده به منظور تایید و یا رد مدل فراهم می‌کند.

۱-۵-۲ مدل‌سازی شناختی اعتیاد

در مدل‌سازی شناختی رفتار عامل در یک آزمون شناختی[2] مشخص ثبت شده و سعی می‌شود بر اساس روابط ریاضی، رفتار عامل توضیح داده شود. روابط ریاضی باید مبتنی بر توانایی‌های شناختی مانند حافظه، توجه و غیره باشند. در این صورت امکان ارتباط یک نوع رفتار مشاهده‌شده با زیر-

[1] Impulsivity
[2] Cognitive assessment task

ساخت‌های روانی آن فراهم می‌شود. در بخش ۵-۱، این نوع مدل‌سازی و تفاوت آن با دیگر روش‌های پیشین توضیح داده شده است.

۱-۶ ساختار پایان‌نامه

همان‌طور که بیان شد، پایان‌نامه‌ی حاضر در دو بخش تنظیم شده است. در بخش اول سعی شده است بر اساس روش‌شناسیِ بیان‌شده در بخش قبل، یک مدل عصبی-محاسباتی برای اعتیاد، که نسبت به مدل‌های پیشین انطباق بیشتری با داده‌های رفتاری و ساختاری دارد، ارائه شود. بدین منظور، در فصل ۲ مروری بر مدل‌های عصبی-محاسباتیِ تصمیم‌گیری شده است. مدل‌های بیان شده بر اساس یادگیری تقویتی استوار هستند، که پایه‌ی مدل پیشنهاد شده در این پایان‌نامه است. در فصل ۳ مدل‌های پیشین اعتیاد به مواد مرور شده‌اند و زیربنای عصبی، نقاط ضعف و قوت آنها برشمرده شده است. مدل‌های بیان‌شده در این بخش به دو دسته‌ی مدل‌های مبتنی بر عصب‌شناسی محاسباتی و دارویی تقسیم‌بندی شده‌اند. یکی از مدل‌های بیان‌شده در این بخش مبنای مدل پیشنهادشده در این پژوهش را تشکیل می‌دهد. در فصل ۴ به معرفی مدل پیشنهادشده در پژوهش می‌پردازیم و در مورد جنبه‌های مختلف آن بحث می‌کنیم. در بخش دوم، که شامل فصل ۵ است، مدل‌سازی شناختی آزمون قمار آیوا بر اساس داده‌های مراجعه کنندگان به *مرکز ملی مطالعات اعتیاد ایران* بیان شده است. در انتها در فصل ۶ به ارائه‌ی خلاصه‌ی پایان‌نامه و پیشنهاد کارهای قابل انجام در ادامه این پژوهش می‌پردازیم.

بخش اول:
مدل‌سازیِ عصبی-محاسباتی اعتیاد

فصل ۲

یادگیری تقویتی، شرطی شدن و نظام دوپامینی

۲-۱ مقدمه

نشانه‌های رفتاریِ اعتیاد که در بخش قبل ذکر شدند، در حالت کلی یک آسیبِ نظام تصمیم‌گیری و انتخاب به شمار می‌روند. همانطور که بیان شد، ارائه‌ی یک مدلِ عصبی-محاسباتی برای اعتیاد، در ابتدا نیازمند معرفی یک مدل عصبی-محاسباتی از تصمیم‌گیری در یک فرد سالم (بدون سابقه‌ی مصرف مواد) است. در پژوهش حاضر، بر گونه‌ی خاصی از مدل‌های تصمیم‌گیری که مدل‌های مبتنی بر ارزش[1] نامیده می‌شوند، تمرکز شده است [۱۶]. منطق تصمیم‌گیری در این مدل‌ها بدین‌صورت است که حیوان بر اساس ارزش نسبی که به عمل‌های مختلف نسبت داده است، در یک وضعیت مشخص عملی را انتخاب می‌کند. مدل یادگیری تقویتی، یکی از اعضای این خانواده از مدل‌های تصمیم‌گیری است و دارای این ویژگی مطلوب است که زیربنایِ روانی و عصبی آن به طور گسترده مطالعه شده است. این مدل به لحاظِ روانی، توضیح‌دهنده‌ی آزمایش‌های شرطی شدن است و به لحاظ عصبی انطباق مطلوبی با ساختار نظامِ دوپامینیِ مغز دارد. به این علت، یادگیری تقویتی مبنای مدل معرفی‌شده در این پژوهش قرار داده شده است. در بخش بعد، مدلِ یادگیری تقویتی و گونه‌های

[1] Value-based

مختلف آن را شرح می‌دهیم. در بخش ۲-۳ به توضیح ارتباط این مدل با داده‌های رفتاری حاصل از شرطی شدن می‌پردازیم. در انتها نیز، زیرساخت‌های عصبی این مدل و تناظر بین مولفه‌های آن و بخش‌های مختلف نظام دوپامینی ارائه می‌شود.

۲-۲ یادگیری تقویتی

یادگیری تقویتی [۱۷] به حل مسئله‌ی کنترل بهینه در یک محیط با عدم قطعیت می‌پردازد. هدف عامل تصمیم‌گیرنده، انتخاب اعمالی[۱] است که شاخصی از پاداش[۲] را بهینه کند. محیط تصمیم‌گیری به صورت یک فرایند تصمیم‌گیری مارکوف[۳] مدل‌سازی می‌شود. یک محیط مارکوف به وسیله‌ی دو تابع R و T که بر روی مجموعه $\mathcal{S}$ از وضعیت‌ها[۴] و مجموعه $\mathcal{A}$ از عمل‌ها تعریف شده‌اند، مشخص می‌شود. در هر لحظه از زمان مانند t، وضعیت محیط به وسیله متغیر تصادفی s_t ($s_t \in \mathcal{S}$) نشان داده می‌شود. با انتخاب عمل a_t ($a_t \in \mathcal{A}$) توسط عامل، محیط در واحد زمانی $t+1$ وارد وضعیت s_{t+1} می‌شود. با معلوم بودن s_t و a_t، T که تابع انتقال[۵] نام دارد، تابع توزیعِ احتمالِ وضعیت بعدی، s_{t+1}، را مشخص می‌کند. این تابع توزیع را به صورت زیر نشان می‌دهیم:

$$\mathrm{T}^{a}_{ss'} = P(s_{t+1} = s' | s_t = s, a_t = a)$$

در هر واحد زمانی t پس از انتخاب عمل a_t، عامل پاداش r_t را از محیط دریافت می‌کند. این پاداش دریافتی به‌وسیله‌ی تابع پاداش R تعیین می‌شود و توزیعِ احتمال آن را به صورت زیر نشان می‌دهیم:

$$\mathrm{R}^{a}_{s,r} = P(r_t = r | s_t = s, a_t = a)$$

مارکوف بودن یک محیط به این خاصیت اشاره می‌کند که وضعیت بعدیِ محیط تنها به وضعیت کنونی آن و عمل فعلی عامل بستگی دارد و وضعیت‌های قبلی و عمل‌های قبلی عامل تاثیری بر تحولات بعدی محیط ندارد.

---

[1] Actions
[2] Reward
[3] Markov decision process (MDP)
[4] States
[5] Transition function

یک سیاست مارکوف[1]، یا به طور خلاصه یک سیاست، به صورت یک نگاشت از فضای وضعیت‌ها به فضای عمل‌ها تعریف می‌شود. سیاست انتخاب عمل می‌تواند قطعی[2] یا غیرقطعی[3] باشد. در صورتی که انتخاب عمل به صورت غیرقطعی باشد، احتمال انتخاب عمل a در وضعیت s را با نماد $\pi_{s,a}$ نشان می‌دهیم:

$$\pi_{s,a} = P(s_t = s, a_t = a)$$

اگر عامل از وضعیت s شروع کند و بر اساس یک سیاست مشخص مانند π انتخاب عمل کند، آن‌گاه می‌توان فرض کرد که محیط تحت سیاست عامل از وضعیتی به وضعیت دیگری منتقل می‌شود. با این نگاه می‌توان توابعِ انتقال و پاداش را برای یک سیاست مانند π به صورت زیر تعریف کرد:

$$\mathrm{T}^{\pi}_{ss'} = \sum_{a \in \mathcal{A}} \pi_{s,a} . \mathrm{T}^{a}_{ss'}$$
$$\mathrm{R}^{\pi}_{s,r} = \sum_{a \in \mathcal{A}} \pi_{s,a} . \mathrm{R}^{a}_{s,r}$$

هدف از یادگیری تقویتی یافتن سیاستی است که عامل با تبعیت از آن معیاری از پاداش‌های دریافتیِ خود را بهنیه کند. این معیار مشخص، ارزش یک سیاست نام دارد که شاخصی است از پاداش‌هایی که عامل انتظار دارد با دنبال کردن آن سیاست دریافت کند. ارزش یک سیاست مانند π را با شروع از وضعیت اولیه s با تابع ارزش[4] V^{π}_{s} نشان می‌دهیم. به منظور ساده‌سازی در شرایطی که سیاست انتخاب عمل ثابت است، اندیس π را از تابع ارزش، انتقال و پاداش حذف می‌کنیم. با استفاده از تعریف تابع ارزش می‌توان سیاست بهینه را تعریف کرد. سیاست π^* بهینه است اگر به ازای هر وضعیت s و هر سیاست دلخواه π داشته باشیم $\mathrm{V}^{\pi}_{s} \leq \mathrm{V}^{\pi^*}_{s}$. در بخش بعدی به تعریف سه نوع تابع ارزش متداول که در این پژوهش نیز مورد استفاده قرار گرفته‌اند، می‌پردازیم.

[1] Markov policy
[2] Deterministic
[3] Non-deterministic
[4] Value function

۲-۲-۱ تابع ارزش

به طور شهودی انتظار می‌رود که تابع ارزش بازنمایی کننده میزان کل پاداشی باشد که عامل در طول عملکرد خود دریافت می‌کند. از این ایده می‌توان برای تعریف تابع ارزش در محیط‌هایی که عامل پس از تعداد محدودی عمل وارد یک وضعیت پایانی[1] می‌شود، استفاده کرد. وضعیت پایانی، وضعیتی است که عامل برای همیشه در آن مانده و پاداش صفر دریافت می‌کند. اگر فرض کنیم که عامل پس از T واحد زمانی وارد یک وضعیت پایانی می‌شود، آنگاه تابع ارزش به صورت زیر خواهد بود:

$$\mathrm{V}_{s_t} = E[r_t + r_{t+1} + r_{t+2} + \cdots + r_T | s_t] \qquad (۱-۲)$$

در حالتی که عامل پس از طی تعداد محدودی وضعیت وارد یک وضعیت پایانی نشود، (۲-۱) یک تعریف معتبر برای تابع ارزش نخواهد بود؛ زیرا امکان عدم همگرایی آن وجود دارد. در ادامه‌ی این بخش تابع ارزش‌های دیگری که در محیط‌های بدون وضعیت پایانی قابل استفاده هستند معرفی می‌کنیم.

با بسط (۲-۱) آن را می‌توان به صورت زیر نوشت:

$$\mathrm{V}_{s_t} = E_r[\mathrm{R}_{s_t}] + \sum_{s_{t+1} \in \mathcal{S}} \mathrm{T}_{s_t s_{t+1}} \mathrm{V}_{s_{t+1}} \qquad (۲-۲)$$

صورت کلی رابطه‌ی بازگشتی فوق به معادله بلمن[2] شهرت دارد. معادله‌ی بلمن پایه و اساس یادگیری تقویتی محسوب می‌شود. در واقع روابط بهینگی بلمن وجود جواب بهینه را تضمین و شرایط لازم و کافی برای مقادیر ارزش وضعیت‌ها تحت سیاست بهنیه را بیان می‌کنند. در مورد تابع ارزش (۲-۱)، می‌توان نشان داد که مقادیر ارزش‌ها تحت سیاست بهینه در روابط زیر صدق می‌کنند:

$$\mathrm{V}_s^{\pi^*} = \max_{a \in \mathcal{A}} \left(E_r[\mathrm{R}_s^a] + \sum_{s' \in \mathcal{S}} \mathrm{T}_{ss'}^a \mathrm{V}_{s'}^{\pi^*} \right) \qquad (۳-۲)$$

[1] Absorbing state
[2] Bellman

تعاریف گوناگونی می‌توان برای ارزش یک وضعیت در افق زمانی نامحدود (محیط بدون وضعیت پایانی) در نظر گرفت. به عنوان مثال می‌توان مجموع پاداش‌های دریافت شده در n عمل آخر را به عنوان ارزش یک سیاست در نظر گرفت. شرط محدودکننده این است که برای تعریف بیان‌شده، باید رابطه‌ی بلمن وجود داشته باشد که در نتیجه بتوان با رویکرد یادگیری تقویتی سیاست بهینه را کشف کرد. در اینجا ما به تعریف دو تابع ارزش متداول که در این پژوهش نیز مورد استفاده قرار گرفته‌اند می‌پردازیم: تابع ارزش کاهش نمایی[1] و تابع ارزش متوسط پاداش[2]. این دو تابع ارزش در افق زمانی نامحدود همگرا می‌شوند.

تابع ارزش کاهش نمایی، تقریبی از میانگین پاداش‌های دریافت شده در n عمل آخر انجام شده توسط عامل است. مقدار n به پارامتر نرخ کاهش[3] وابسته است که آن را با γ نشان می‌دهیم ($0 \leq \gamma \leq 1$). تعریف این تابع ارزش به صورت زیر است:

$$\mathrm{V}_{s_t} = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+1} + \ldots | s_t] = E\left[\sum_{\tau \geq t} \gamma^{\tau - t} r_\tau | s_t\right] \qquad (۴-۲)$$

با توجه به (۲-۴)، تاثیر پاداش‌های دریافت‌شده در آینده بر روی ارزش وضعیت s_t به صورت نمایی کاهش می‌یابد و این امر علت نام‌گذاری این تابع ارزش به کاهش نمایی است. معادله‌ی بلمن بازگشتی متناظر با (۲-۴) به صورت زیر است:

$$\mathrm{V}_{s_t} = E_r[\mathrm{R}_{s_t}] + \gamma . E_{s_{t+1}}[\mathrm{V}_{s_{t+1}}] \qquad (۵-۲)$$

همانطور که از (۲-۵) مشخص است، پارامتر نرخ کاهش (γ) تعیین کننده‌ی میزان تاثیر ارزش یک وضعیت بر وضعیت پیشین خود است. مقادیر نزدیک به صفر این پارامتر به معنای تاثیر کم پاداش‌های دریافت شده در آینده در ارزش وضعیت فعلی و مقادیر نزدیک به یک به معنای عدم کاهش ارزش پاداش‌های دریافت شده در آینده است.

معادله‌ی بهینگی بلمن برای تابع ارزش (۲-۲) به صورت زیر است:

[1] Exponential discounting value function
[2] Average-reward value function
[3] Discount factor

$$\mathrm{V}_{s}^{\pi^{*}} = \max_{a \in \mathcal{A}} \left(E_r[\mathrm{R}_{s}^{a}] + \sum_{s' \in \mathcal{S}} \gamma \mathrm{T}_{ss'}^{a} \mathrm{V}_{s'}^{\pi^{*}} \right) \quad (۶-۲)$$

همانطور که ذکر شد در تابع ارزش معرفی‌شده مقادیر پاداش‌های آینده به طور نمایی کاهش می‌یابد. این نوع کاهش ارزش اغلب با رفتار انسان و حیوان که در آنها ارزش پاداش‌های آینده به صورت هیپربولیک[1] کاهش داده می‌شود، ناسازگار است [۱۸]. کاهش هیپربولیک بدان معناست که ارزشِ یک پاداش متناسب با فاصله‌ی زمانی دریافت آن کم می‌شود. بر اساس این منطق می‌توان تابع ارزش را به صورت زیر در نظر گرفت:

$$\mathrm{V}_{s_t} = E\left[\sum_{\tau \geq t} \frac{r_\tau}{\tau} \right] \quad (۷-۲)$$

متاسفانه مقادیر ارزش تعریف شده در (۲-۷) را نمی‌توان به صورت معادله بازگشتی بلمن نوشت و بنابراین قابل استفاده در یادگیری تقویتی نیست. به جای تعریف فوق می‌توان از تعریف دیگری برای مقادیر ارزش‌ها استفاده کرد که در آن تاثیر پاداش‌های آینده شبیه به کاهش هیپربولیک کم می‌شود. در این روش که متوسط پاداش نام دارد [۱۹]، معادله‌ی بازگشتی بلمن به صورت زیر است:

$$\mathrm{V}_{s_t}^{\pi} = E_r[\mathrm{R}_{s_t}^{\pi}] - \rho^{\pi} + E_{s_{t+1}}[\mathrm{V}_{s_{t+1}}^{\pi}] \quad (۸-۲)$$

که در (۲-۸)، ρ^{π} میانگین پاداش‌های دریافت شده تحت سیاست π است. می‌توان نشان داد که در محیط‌های مارکوفی که با شروع از هر وضعیت دلخواه و هر سیاست دلخواه، عامل با احتمال یک از هر وضعیتی می‌گذرد، میزان متوسط پاداش مستقل از وضعیت شروع است. مقدار متوسط پاداش را می‌توان به صورت زیر تعریف کرد:

$$\rho_{s_t}^{\pi} = \lim_{N \to \infty} \frac{E(\sum_{\tau=t}^{N-1} R_{s_\tau}^{\pi})}{N} \quad (۹-۲)$$

همانطور که از تعریف ارائه شده در (۲-۸) به نظر می‌رسد، مقدار ارزش وضعیت بعد (s_{t+1}) بدون کاهش در ارزش وضعیت فعلی شرکت می‌کند. ولی این بدان معنا نیست که ارزش یک وضعیت نسبت

[1] Hyperbolic

به زمان رسیدن پاداش‌ها در آینده حساس نیست. در حقیقت می‌توان رابطه (۲-۸) را به این گونه دید که مقدار ارزش وضعیت‌های بعد ($E_{s_{t+1}}[\mathrm{V}_{s_{t+1}}]$) از متوسط پاداش ($\rho^{\pi}$) کم شده‌اند و حاصل نهایی در ارزش وضعیت فعلی تاثیر گذاشته است. منطق این نوع کاهش ارزش را می‌توان بدین‌گونه بیان کرد که انتقال عامل از وضعیت s_t به وضعیت s_{t+1} در یک واحد زمانی صورت می‌گیرد که در طول آن عامل پاداشی به اندازه‌ی میانگین پاداش در زمان (ρ) از دست می‌دهد و این مقدار از ارزش وضعیت-های بعدی کم شده است. از این بیان در بخش ۴-۴-۴ استفاده می‌شود.

با گسترش (۲-۸)، می‌توان مقدار ارزش هر وضعیت را به صورت زیر نوشت:

$$\mathrm{V}^{\pi}_{s_t} = E\left[\sum_{\tau \geq t}(\mathrm{R}^{\pi}_{s_\tau} - \rho^{\pi})\right] \qquad (۱۰-۲)$$

رابطه‌ی فوق تابع ارزشِ میانگینِ تنظیم‌شده[1] نام دارد و تعبیر دیگری از تابع ارزشِ میانگین پاداش ارائه می‌کند: پاداش‌ها نسبت به متوسط آنها اندازه‌گیری می‌شوند و در تابع ارزش شرکت داده می-شوند. از این تعبیر در بخش ۴-۳ استفاده می‌شود.

معادله‌ی بهینگی بلمن برای تابع ارزش متوسط پاداش به صورت زیر است [۲۰]:

$$\mathrm{V}^{\pi^*}_{s} + \rho^{*} = \max_{a \in \mathcal{A}}\left(E_r[\mathrm{R}^{a}_{s}] + \sum_{s' \in \mathcal{S}} \mathrm{T}^{a}_{ss'}\mathrm{V}^{\pi^*}_{s'}\right) \qquad (۱۱-۲)$$

که در رابطه‌ی فوق ρ^{*}، مقدار میانگین پاداش تحتِ سیاست بهنیه است.

در این بخش به سه روش مختلف برای تعریف تابع ارزش اشاره شد. در بخش بعدی به نحوه‌ی محاسبه مقادیر بهینه‌ی ارزش یک وضعیت و سیاست بهینه می‌پردازیم.

۲-۲-۲ پیش‌بینی پاداش

در هر سه نوع تابع ارزش معرفی شده در بخش قبل می‌توان از حل دستگاه معادلات بازگشتی بلمن برای پیدا کردن ارزش یک سیاست استفاده کرد. همچنین برای پیدا کردن مقادیر بهنیه می‌توان از روابط بهینگی بلمن استفاده کرد. لیکن انجام این کار نیاز به اطلاع کامل از محیط (توابع R و T)

[1] Average-adjusted value function

دارد، که در شرایط واقعی عامل از آنها آگاه نیست. به علاوه در شرایطی که فضای حالت بزرگ است، نمی‌توان معادلات بلمن را به نحو کارا حل کرد. برای حل مشکل اخیر می‌توان از روش تکرار ارزش[۱] استفاده کرد. در این روش، ارزش وضعیت‌ها مقداردهی اولیه می‌شوند و سپس طبق رابطه زیر به‌روز می‌گردند:

$$\hat{V}_s(n+1) \leftarrow E_r[\mathrm{R}_s] + \sum_{s' \in \mathcal{S}} \mathrm{T}_{ss'} \hat{\mathrm{V}}_{s'}(n), \forall s \in \mathcal{S} \qquad (۱۲-۲)$$

در حالت کلی به روابطی که از دانش R و T در تخمین استفاده می‌کنند رویکردهای مبتنی بر مدل[۲] اطلاق می‌شود. در مقابل این رویکرد، روش‌های مستقل از مدل[۳] قرار دارد. تئوری‌های دوپامینی نظام عصبی بر رویکرد نوع دوم بنا شده‌اند.

در رویکردهای مستقل از مدل، عامل بر اساس نمونه‌هایی که از پاداش و وضعیت خود در محیط دریافت می‌کند تابع ارزشِ سیاست خود را تخمین می‌زند. برای تخمین تابع ارزش روش‌های مختلفی وجود دارد. به عنوان مثال، عامل می‌تواند با استفاده از تجربه‌های خود مدل محیط را یادگرفته و سپس با استفاده از روش بیان شده در بالا ارزش وضعیت‌ها را تخمین بزند. این امکان نیز وجود دارد که ارزش وضعیت‌ها به طور مستقیم از روی نمونه‌های مشاهده شده بدون واسطه مدل محیط یاد گرفته شوند. روشی که در اینجا بیان می‌کنیم بر اساس رویکرد دوم استوار است. فرض کنیم که عامل پاداش r_t را مشاهده کند و از وضعیت s_t به وضعیت s_{t+1} منتقل شود. در این صورت می‌توان r_t را به عنوان نمونه‌ای از $E_r[\mathrm{R}_{s_t}]$ انگاشت. به همین نحو می‌توان تخمین فعلی از ارزشِ وضعیت s_{t+1} که برابر $\hat{\mathrm{V}}_{s_{t+1}}$ است را به عنوان نمونه‌ای از $E_{s_{t+1}}[\mathrm{V}_{s_{t+1}}]$ در نظر گرفت. با این فرض می‌توان با استفاده از روابط بازگشتی بلمن نمونه‌ای از ارزش وضعیت فعلی ارائه کرد. به عنوان مثال در مورد تابع ارزش کاهش نمایی، با استفاده از (۲-۵)، $r_t + \gamma \hat{\mathrm{V}}_{s_{t+1}}$ نمونه‌ای از ارزش وضعیت فعلی خواهد بود. اختلاف

[۱] Value iteration
[۲] Model-based
[3] Model-free

این نمونه از ارزش وضعیت فعلی با تخمین وضعیت فعلی *خطای تفاضل زمانی*[1] نام دارد که آن را با δ_t نشان می‌دهیم:

$$\delta_t = r_t + \gamma \hat{\mathrm{V}}_{s_{t+1}} - \hat{\mathrm{V}}_{s_t} \qquad (۱۳-۲)$$

اگر مقدار تخمین از ارزش وضعیت s_t را در جهت کم کردن میزان خطا به‌روز کنیم خواهیم داشت:

$$\hat{\mathrm{V}}_{s_t} \leftarrow \hat{\mathrm{V}}_{s_t} + \alpha \delta_t \qquad (۱۴-۲)$$

که در رابطه فوق α ($0 \leq \alpha \leq 1$) نرخ یادگیری نامیده می‌شود و مقدار آن تعیین می‌کند که تجربیات جدید تا چه حد ارزش وضعیت‌ها را تحت تاثیر قرار دهند. با اعمال شرایطی بر α و فرایند مارکوف، روش یادگیری فوق همگرا می‌شود. به اینگونه یادگیری که بر اساس سیگنال خطای تفاضل زمانی است، یادگیری تقویتی تفاضل زمانی[2] اطلاق می‌شود [۲۱] و هسته‌ی اصلی مدل استفاده‌شده در این پژوهش را تشکیل می‌دهد.

در مورد تابع ارزش میانگین پاداش، با استفاده از (۲-۸)، خطای تفاضل زمانی [۱۹] به صورت زیر خواهد بود:

$$\delta_t = r_t - \rho_t + \hat{\mathrm{V}}_{s_{t+1}} - \hat{\mathrm{V}}_{s_t} \qquad (۱۵-۲)$$

که در رابطه بالا چون تخمین ارزش‌ها به صورت برخط انجام می‌گیرد، مقدار متوسط پاداش نیز باید تخمین زده شود. این کار را می‌توان به وسیله میانگین‌گیری از پاداش‌های دریافت شده انجام داد. در اینجا ما از روش میانگین‌گیری پنجره‌ای[3] استفاده می‌کنیم:

$$\rho_{t+1} \leftarrow (1-\sigma)\rho_t + \sigma r_t \qquad (۱۶-۲)$$

[1] Temporal difference error
[2] Temporal-difference reinforcement learning (TDRL)
[3] Moving average

که در رابطه فوق σ به طور تقریبی اندازه‌ی پنجره‌ی میانگین‌گیری را تعیین می‌کند ($\sigma \ll \alpha$). شرایط همگرایی تخمین ارزش‌ها، در تابع ارزش میانگین پاداش‌ها وابسته به شرایط محیط مارکوف است که در [۲۰] بحث شده است.

به عنوان تعمیمی از فرایند مارکوف بیان شده در قبل، می‌توان مسئله را شبه‌مارکوف[1] فرض کرد. در محیطِ شبه‌مارکوف در لحظه t، انتقال از یک وضعیت به وضعیت دیگر در زمان d_t صورت می‌گیرد. در این صورت سیگنال خطا در مورد تابع ارزش کاهش نمایی به صورت زیر خواهد بود:

$$\delta_t = r_t + \gamma^{d_t} \hat{\mathrm{V}}_{s_{t+1}} - \hat{\mathrm{V}}_{s_t} \qquad (۱۷-۲)$$

و در مورد تابع ارزش متوسط پاداش به صورت زیر:

$$\delta_t = r_t - d_t \rho_t + \hat{\mathrm{V}}_{s_{t+1}} - \hat{\mathrm{V}}_{s_t} \qquad (۱۸-۲)$$

که توجیه رابطه‌ی فوق همانند قبل خواهد بود. می‌توان تصور کرد که با اتلاف d_t واحد زمانی ، عامل پاداش به اندازه‌ی $d_t \rho_t$ از دست می‌دهد، که این مقدار از ارزش وضعیت بعد کم شده است.

۲-۲-۳ سیاست بهینه

با داشتن مقادیر بهینه ارزش هر وضعیت ($\mathrm{V}^{\pi^*}_{s_t}$)، که از معادلات بهینگی بلمن نتیجه می‌شوند، می‌توان سیاست بهینه را تعیین کرد:

$$\pi^*_s = \arg_{a \in \mathcal{A}} \max \left(E_r[\mathrm{R}^a_s] + \sum_{s' \in \mathcal{S}} \mathrm{T}^a_{ss'} \mathrm{V}^{\pi^*}_{s'} \right) \qquad (۱۹-۲)$$

در صورت مشخص نبودن مقادیر بهینه، می‌توان از یک سیاست اولیه شروع کرده و به تدریج آن را بهبود داد. به این منظور سیاست π را در نظر گرفته و سیاست π' را به صورت زیر تعریف می‌کنیم:

$$\pi'_s = \arg_{a \in \mathcal{A}} \max \left(E_r[\mathrm{R}^a_s] + \sum_{s' \in \mathcal{S}} \mathrm{T}^a_{ss'} \mathrm{V}^{\pi}_{s'} \right) \qquad (۲۰-۲)$$

[1] Semi-Markov

که در این صورت سیاست π' از سیاست π بدتر نیست. با تکرار رابطه فوق به تدریج می‌توان سیاست اولیه را بهبود بخشید تا در نهایت به سمت سیاست بهینه همگرا شود.

روش فوق قادر به پیدا کردن سیاست بهینه در رویکرد مستقل از مدل مانند روش یادگیری تفاضل زمانی بیان شده در بخش پیشین نیست. برای حل این مشکل و یادگیری سیاست به صورت برخط، می‌توان سیاست بهینه را برحسب یک بردار پارامتر بیان کرده، در طول یادگیری با روش بیشترین شیب[1] مقدار بردار پارامتر را در جهت بهتر شدن سیاست تغییر داد. روش دیگری که در اینجا بیان می‌کنیم و در این پژوهش مورد استفاده قرار گرفته است، بر اساس تخمین ارزش یک زوج وضعیت-عمل[2] استوار است. در این روش به جای محاسبه‌ی ارزش یک وضعیت، ارزش انتخاب یک عمل مشخص در آن وضعیت تخمین زده می‌شود. ارزش انتخاب عمل a در وضعیت s با Q_s^a نشان داده شده و مطابق زیر تعریف می‌شود:

$$\mathrm{Q}_s^a = E[\mathrm{R}_s^a] + \sum_{s' \in \mathcal{S}} \mathrm{T}_{s,s'}^a . \max_{a' \in \mathcal{A}} \mathrm{Q}_{s'}^{a'} \qquad (۲۱-۲)$$

بر خلاف مقادیر ارزش وضعیت‌ها، مقادیر جفت‌های وضعیت-عمل مستقل از سیاست انتخاب عمل هستند. همانند مقادیر ارزش وضعیت‌ها می‌توان مقادیر ارزش وضعیت-عمل‌ها را به وسیله یادگیری تفاضل زمانی به روز کرده و به سمت مقادیر بهینه همگرا شد [۲۲]. به عنوان مثال در مورد تابع ارزش کاهشی نمایی، خطای تفاضل زمانی به صورت زیر خواهد بود:

$$\delta_t = r_t + \gamma \mathrm{V}_{s_{t+1}} - \mathrm{Q}_{s_t}^a \qquad (۲۲-۲)$$

که در رابطه‌ی فوق $\mathrm{V}_{s_t} = \max_{a \in \mathcal{A}} \mathrm{Q}_{s_{t+1}}^a$ و یادگیری بر اساس رابطه فوق Q-یادگیری نامیده می‌شود. در صورتی که در (۲-۲۲) مقدار V_{s_t} برابر $\mathrm{Q}_{s_t}^{\pi}$ در نظر گرفته شود یادگیری سارسا[3] نامیده می‌شود. با توجه به (۲-۲۲) می‌توان از روی مقادیر ارزش وضعیت-عمل‌ها سیاست بهینه را کشف کرد. بدین صورت که در هر وضعیت، عملی انتخاب شود که بیشترین مقدار ارزشِ وضعیت-عمل را دارد:

[1] Gradient descent
[2] State-action pair
[3] State-Action-Reward-State-Action (SARSA)

$$\pi^*_{s,\arg_{a\in\mathcal{A}}\max Q^a_s} = 1 \qquad (۲۳-۲)$$

در صورتی که انتخاب عمل‌ها به صورت برخط انجام گیرد، و در نتیجه تخمین ارزش‌ها همزمان با تشکیلِ سیاستِ بهینه باشد، از (۲-۲۳) نمی‌توان استفاده کرد و باید گرایش به سمت رفتار بهینه همزمان با تخمین صحیح ارزش جفت‌های وضعیت-عمل باشد. در بخش بعد به چگونگی این کار می-پردازیم.

۲-۲-۴ انتخاب عمل

مسئله‌ی اصلی در انتخاب عمل برخط، برقراری تعادل بین بهره‌برداری از اطلاعات پیشین و به‌دست آوردن اطلاعات جدید از محیط است. این مسئله که با نام تعادل اکتشاف و بهره‌برداری[۱] شناخته شده است، به این نکته اشاره می‌کند که اگر عامل قبل از بدست آوردن تخمین قابل قبول از ارزش وضعیت‌ها به سمت سیاست خاصی گرایش پیدا کند، ممکن است که در طول حیات خود بر اساس سیاست غیربهنیه رفتار کند. از طرف دیگر اگر عامل در طول حیات خود تنها به جستجوی محیط پرداخته و از اطلاعات پیشین خود درباره‌ی ارزش وضعیت‌ها استفاده نکند، کارایی پایینی در طول وظیفه خود خواهد داشت. برای حل این مشکل و برقراری تعادل میان اکتشاف و بهره‌برداری در اینجا دو راه حل بیان می‌شود.

در راه حل اول که ϵ-حریصانه[۲] نام دارد، عامل با احتمال ϵ عملِ با بیشترین ارزش پیش‌بینی شده را انتخاب کرده و با احتمال $1-\epsilon$، یک عمل تصادفی انتخاب می‌کند. بنابراین در این روش عامل در طول عمر خود میزانی از انتخاب‌های خود را به اکتشاف محیط اختصاص می‌دهد. این روش انتخاب عمل روشی است که در الگوریتم‌های یادگیری مبتنی بر میانگین پاداش استفاده می‌شود. در این الگوریتم‌ها (از جمله الگوریتم یادگیری تفاضلی زمانی که در بخش پیش بیان شد)، میانگین پاداش، ρ_t ، بر روی عمل‌های غیر اکتشافی محاسبه می‌شود. به این معنا که:

[۱] Exploration-exploitation balance
[۲] ϵ-greedy

$$\rho_{t+1} = \begin{cases} (1-\sigma)\rho_t + \sigma r_t & \text{if } a_t \text{ is not exploratory} \\ \rho_t & \text{otherwise} \end{cases} \quad (۲۴-۲)$$

در راه دوم که بیشینه نرم[1] نامیده می‌شود، عامل در وضعیت s عمل a را با احتمالی متناسب با ارزش آن انتخاب می‌کند:

$$\pi_{s,a} = \frac{e^{Q_s^a/\beta}}{\sum_{a' \in \mathcal{A}} e^{Q_s^{a'}/\beta}} \quad (۲۵-۲)$$

که در رابطه فوق β تعیین‌کننده‌ی میزان متناسب بودن احتمال انتخاب عمل با ارزش یک عمل است. در بخش بعدی با توجه به نمادگذاری و الگوریتم‌های یادگیری بیان شده، به توانایی توصیف مدل یادگیری تقویتی برای توصیف داده‌های رفتاری می‌پردازیم.

۲-۳ شرطی شدن

۲-۳-۱ داده‌های رفتاری

از دیرباز چگونگی تصمیم‌گیری در حیوان‌ها در حوزه روان‌شناسیِ رفتاری مطالعه شده است. در این حوزه رفتار حیوان‌ها در برابر یک محرک[2] مشخص، هنگامی که پاسخ‌های حیوان با پاداش و یا تنبیه[3] همراه است مطالعه می‌شود. یک محرک می‌تواند یک لامپ که حیوان آن را مشاهده می‌کند و یا شنیدن صدای بوق توسط حیوان باشد. با استفاده از نمادگذاری معرفی شده در بخش قبل، می‌توان مشاهده‌ی یک محرک توسط حیوان را به معنای قرار گرفتن حیوان در یک وضعیت مشخص تلقی کرد.

آزمایش‌های شرطی شدن به این‌گونه هستند که بین وضعیت حیوان، عملکرد حیوان در آن وضعیت و پاداش دریافت‌شده توسط او ارتباط مشخصی برقرار می‌شود. منظور از عملکرد حیوان انجام عمل مشخصی مانند فشار دادن یک اهرم، حرکت کردن به سمت مشخصی و غیره است. مطابق با نمادگذاری بخش پیش، عمل حیوان در لحظه t را با a_t نشان می‌دهیم. پاداش دریافت شده می‌تواند

[1] Soft-max
[2] Stimulus
[3] Punishment

جدول ۲-۱ برخی از آزمایش‌های شرطی شدن کلاسیک

پدیده	مرحله اول	مرحله دوم	آزمایش	منبع
فراگیری	A→US		A⇒CR	[۲۳]
ترک رفتار	A→US	A→-	A⇒-	[۲۳]
بلوکه کردن	A→US	A,B→US	B⇒-	[۲۴]
شرطی شدن مرحله دوم	A→US	A,B→-	B⇒CR	[۲۵]

مثلا دادن محلولی از ساکروز به موش و تنبیه می‌تواند دادن شوک الکتریکی به آن باشد، که آن را با r_t نشان می‌دهیم.

به طور سنتی آزمایش‌های شرطی شدن به دو حوزه‌ی شرطی شدن کلاسیک[۱] (و یا پاولوفی[۲]) و شرطی شدن ابزاری[۳] تقسیم‌بندی می‌شود. در شرطی شدن کلاسیک، حیوان مستقل از عمل خود پاداش و/یا تنبیه دریافت می‌کند؛ بنابراین هدف او تنها پیش‌بینی پاداش دریافت شده پس از مشاهده یک محرک است. در شرطی شدن ابزاری، پاداش یا تنبیه‌ای که حیوان دریافت می‌کند وابسته به عمل او است. بنابراین در این نوع شرطی شدن، حیوان باید ارتباط بین مشاهده‌ی یک محرک، عمل خود و پاداش دریافتی را یاد بگیرد.

در هر یک از انواع شرطی شدن، مجموعه‌ی شواهد رفتاری وسیعی از عملکرد حیوان در شرایط مختلف وجود دارد که برخی از آنها را که در این پژوهش به آنها ارجاع شده است، توضیح می‌دهیم (جدول ۲-۱).

فراگیری[۴] [۲۳] بدان معنی است که هنگامی که یک محرک شرطی‌نشده[۵] (محرک شرطی‌نشده محرکی است که حیوان به طور غریزی به آن ارزش مثبت می‌دهد مانند مزه شیرین ساکروز)، با یک محرک خنثی (مانند مشاهده چراغ)، مرتبط شود (مثلا قبل از دادن ساکروز چراغ روشن شود)، حیوان در صورت دیدن محرک خنثی از خود رفتار شرطی‌شده[۶] نشان می‌دهد (مانند ترشح بزاق دهان).

[1] Classical conditioning
[2] Pavlovian conditioning
[3] Instrumental conditioning
[4] Acquisition
[5] Unconditioned stimulus (US)
[6] Conditioned response (CR)

ترک رفتار[1] [۲۳] به این معنا است که اگر محرکی در مرحله‌ی اول با یک محرک شرطی‌نشده، مرتبط شود (مثلا چراغ با دادن ساکروز همراه شود)، و سپس محرک بدون همراهی محرک شرطی‌نشده ارائه شود (چراغ بدون دادن ساکروز)، در این صورت حیوان با مشاهده‌ی محرک از خود پاسخ شرطی‌شده نشان نمی‌دهد.

پدیده‌ی بلوکه کردن[2] [۲۴] دارای سه مرحله است. در مرحله‌ی اول محرکی مانند A با یک محرک شرطی‌نشده (مثلا ساکروز) همراه می‌شود (مثلا پس از مشاهده‌ی چراغ به حیوان ساکروز داده می‌شود). در مرحله‌ی بعد محرک A همراه با محرک دیگری مانند B، با محرک شرطی‌نشده مرتبط می‌شوند (حیوان چراغ روشن را مشاهده می‌کند و صداق بوق را می‌شنود، سپس ساکروز دریافت می‌کند). در مرحله‌ی آزمایش، حیوان با مشاهده‌ی محرک B از خود پاسخ شرطی‌نشده نشان نمی‌دهد (هنگام شنیدن صدای بوق، بزاق دهنش ترشح نمی‌شود). این امر نشان می‌دهد که محرک A از شرطی شدن محرک B جلوگیری کرده است و یا به اصطلاح آن را بلوکه کرده است.

شرطی شدن مرحله‌ی دوم[3] [۲۵] به این پدیده اشاره می‌کند که اگر یک محرک مانند A با یک محرک شرطی‌نشده همراه شود و سپس محرکی مانند B با محرک A همراه شود، در این صورت حضور محرک B موجب پاسخ شرطی حیوان می‌شود که نشان می‌دهد محرک B با محرک شرطی‌نشده مرتبط شده است.

۲-۳-۲ مدل‌سازی

تئوری‌های مختلفی جهت توضیح انواع شرطی شدن توسعه یافته‌اند که می‌توان آنها را به دو دسته‌ی کلی تقسیم کرد. دسته‌ی اول از مدل‌ها بر پایه مدل پیشنهاد شده در [۲۶] استوار هستند. اینگونه تئوری‌ها از بین خصوصیت‌های بیان شده در جدول ۲-۱ به جز مورد آخر قادر به توضیح همه هستند. اساس کار این مدل‌ها تخمین پاداش آنیِ دریافت شده پس از مشاهده‌ی یک محرک است. از آنجایی که این مدل‌ها به پاداش دریافت شده پس از انجام یک عمل توجه کرده و به پاداش‌های دریافت شده

[1] Extinction
[2] Blocking phenomenon
[3] Second-order conditioning

در مراحل بعد توجهی ندارند، قادر به توضیح شرطی شدن مرحله‌ی دوم نیستند. زیرا اساس شرطی شدن مرحله‌ی دوم ایجاد ارتباط بین دو محرک شرطی‌شده، و در ادبیات یادگیری تقویتی دو وضعیت متوالی، است.

دسته‌ی دوم از مدل‌ها مبتنی بر یادگیری تفاضل زمانی هستند که در این پژوهش نیز از آنها استفاده شده است. در ادامه به بیان اینکه یادگیری تفاضل زمانی هریک از پدیده‌های معرفی شده در جدول ۲-۱ را چگونه توضیح می‌دهد، می‌پردازیم.

فراگیری و ترک رفتار در مدل یادگیری تفاضل زمانی بسیار ساده توضیح داده می‌شود. به خاطر داریم که حضور یک محرک به معنای حضور عامل در یک وضعیت خاص است. مثلا عامل پس از دیدن محرک A در وضعیت S_A است. با این وصف در طول یادگیری ارزش وضعیت S_A، V_{S_A}، به سمت مقدار مورد انتظار پاداش ($E_r[\mathrm{R}_{S_A}]$) میل می‌کند. در نتیجه عامل پس از مشاهده محرک، خود را برای پاداشی که پیش‌بینی می‌کند آماده کرده و پاسخ شرطی‌شده از خود نشان می‌دهد. ترک رفتار نیز به همین نحو توضیح داده می‌شود. اگر پس از یک وضعیت خاص پاداشی به عامل داده نشود، مقدار ارزش وضعیت S_A به سمت صفر همگرا شده، پس از مدتی عامل دیگر با دیدن آن محرک انتظار پاداش ندارد و بنابراین پاسخ شرطی‌شده از خود نشان نمی‌دهد.

توضیح پدیده‌ی شرطی شدن نوع دوم نیز به این صورت است که فرض کنیم عامل در مرحله‌ی اول ارزش وضعیت S_A را که همان $\mathrm{V}_{\mathrm{S_A}}$ است را یاد بگیرد. در مرحله‌ی دوم، پس از وضعیت S_A وارد وضعیت دیگری شود که آن را با S_B نشان می‌دهیم. در این صورت مقدار ارزش وضعیت S_B پس از یادگیری به صورت خواهد بود:

$$\mathrm{V}_{S_B} = E_r[\mathrm{R}_{S_B}] + \gamma \mathrm{V}_{S_A} \qquad (۲۶-۲)$$

بنابراین ارزش پاداش دریافت شده پس از وضعیت S_A به وضعیت S_B نیز منتقل شده است و عامل پس از دیدن محرک B از خود پاسخ شرطی‌شده نشان می‌دهد.

برای توضیح پدیده‌ی بلوکه کردن نیاز به تعمیم رابطه‌ی میان مشاهده‌ی محرک و حضور در یک وضعیت است. این نیاز به این علت است که در مرحله‌ی دوم بلوکه کردن، حیوان به طور همزمان با دو محرک روبرو می‌شود (مانند اینکه هم محرک صدای بوق را بشنود و هم محرک لامپ روشن مشاهده کند) و بنابراین وضعیت حیوان باید ترکیبی از دو محرک باشد. برای بازنمایی چندین محرک در یک وضعیت می‌توان از بردار وضعیت، که آن را با $\mathbf{s}$ نشان می‌دهیم، استفاده کرد. تعریف این بردار به این صورت است که عنصر iام آن در لحظه‌ی t برابر یک است اگر در لحظه t محرک iام حضور داشته باشد؛ و در غیر این صورت صفر است. به این گونه بازنمایی وضعیت، دودویی گفته می‌شود. با این تعریف، ارزش وضعیت در لحظه‌ی t به صورت زیر خواهد بود:

$$\hat{\mathrm{V}}_t = \mathbf{w_t}.\mathbf{s_t} \qquad (۲۷-۲)$$

به منظور یادگیری ارزش هر محرک بردار $\mathbf{w_t}$ توسط سیگنال خطا به روز می‌شود:

$$\mathbf{w_{t+1}} \leftarrow \mathbf{w_t} + \alpha\delta_\mathbf{t}\mathbf{s_t} \qquad (۲۸-۲)$$

سیگنال خطا در (۲-۲۸) همانند قبل محاسبه می‌شود. روش فوق در حالتی که در هر لحظه از زمان فقط یک محرک حضور داشته باشد به مانند روش‌های قبلی عمل می‌کند.

حال پس از معرفی این روش به توضیح پدیده بلوکه کردن می‌پردازیم. اساس توضیح به طور شهودی این است که هنگامی که یک محرک پیشاپیش وقوع یک پاداش را طور کامل پیش‌بینی می‌کند، محرک دیگری که اکنون به طور همزمان وجود دارد با پاداش مرتبط نمی‌شود. بنابراین حیوان در صورت مشاهده‌ی آن به تنهایی از خود پاسخ نشان نمی‌دهد. به لحاظ محاسباتی در مرحله‌ی اول که تنها محرک A حضور دارد بردار وضعیت، $\mathbf{s}$، به صورت $[1,0]^T$ خواهد بود؛ زیرا محرک اول حضور دارد و محرک دوم حضور ندارد. در انتهای یادگیریِ این مرحله، مقدار سینگال خطا به سمت صفر میل کرده و بنابراین ارزش وضعیت به مقدار مورد انتظار پاداش همگرا خواهد شد و در نتیجه بردار وزن $\mathbf{w}$ به صورت $[E[r_t],0]^T$ خواهد بود (r_t پاداش داده‌شده به حیوان پس از مشاهده‌ی محرک A است). در مرحله‌ی بعد، هم محرک A و هم محرک B حضور دارند و بنابراین بردار وضعیت به صورت $[1,1]^T$ خواهد بود. از آنجایی مقدار ضرب داخلی بردار وضعیت در بردار وزن‌ها برابر $E[r_t]$ خواهد بود، مقدار

سیگنال خطا برابر صفر بوده و بنابراین وزن متناظر با محرک B به‌روز نخواهد شد. در مرحله سوم که تنها محرک B حضور دارد، بردار وضعیت به صورت $[0,1]^T$ که با توجه به مقدار بردار وزن که برابر $[E[r_t],0]^T$ است، ارزش پیش‌بینی شده برای محرک B برابر صفر بوده و حیوان از خود پاسخ شرطی شده نشان نمی‌دهد.

همگام با استفاده از مدل یادگیری تفاضل زمانی برای توضیح رفتارهای شرطی شدن، زیرساخت‌های عصبی این مدل نیز شناسایی شده است که در ادامه به توضیح آنها می‌پردازیم.

۲-۴ نظام دوپامینی

توانایی مدل یادگیری تفاضل زمانی در توضیح رفتار حیوانات این سوال را در ذهن ایجاد می‌کند که آیا این مدل در سطح عصبی نیز قادر به توضیح فعالیت‌های نظام عصبی هست یا نه؟ به این معنا که آیا مکانیزم تصمیم‌گیریِ حیوان‌ها در سطح عصبی به وسیله‌ی محاسباتی شبیه محاسبه‌های مدل یادگیری تفاضل زمانی پیاده‌سازی شده است یا خیر؟ در این بخش در پی پاسخ به این پرسش هستیم. در ابتدا به معرفی نظام دوپامینی پرداخته و سپس ارتباط آن را با مدل یادگیری تفاضل زمانی بیان می‌کنیم.

۲-۴-۱ عصب‌شناسی

دوپامین یکی از ناقل‌های عصبی[۱] است که توسط تعداد کمی از نورون‌ها حمل می‌شود. ناقل‌های عصبی موادی شیمیایی هستند که ارتباط بین یک نورون و یک سلول دیگر را ممکن می‌سازند. ناقل-های عصبی در بسته‌هایی در فضایی پیش‌سیناپسی قرار دارند و با آزاد شدن در فضای بین‌سیناپسی توسط گیرنده‌های فضای پس‌سیناپسی جذب می‌شوند.

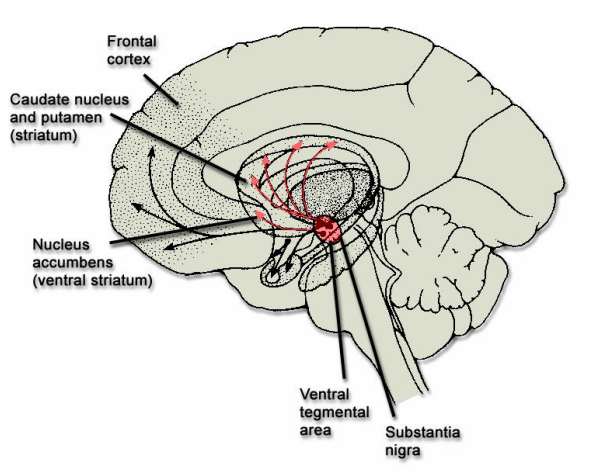
نورون‌هایی که ناقل عصبیِ اصلی در آنها دوپامین است، نورون‌های دوپامینی نامیده می‌شوند. نورون-های حامل دوپامین به طور عمده از ناحیه ونترال تگمنتال[۲] (که آن را با VTA نشان می‌دهیم) و

[۱] Neuromodulator
[۲] Ventral tegmental area (VTA)

ناحیه‌ای موسوم به جسم سیاه[1] (که آن را با SNc نشان می‌دهیم) نشات می‌گیرند. نورون‌های دوپامینی نشات‌گرفته از این نواحی سپس به بخش‌های مختلفی از مغز مانند بخش شکمی استریاتوم[2]، نوکلئوس اکومبنس[3] (که آن را با NA نشان می‌دهیم) و بخش قشری پیش جلویی[4] (که آن را با PFC نشان می‌دهیم) نگاشته می‌شوند (شکل ۲-۱) [۲۷]. نورون‌های دوپامینی در فعالیت‌های زیادی مانند مانند حافظه، کنترل موتور، یادگیری و توجه نقش دارند. همچنین اختلال در عمکرد این نورون‌ها باعث بیماری‌هایی مانند پارکینسون[5]، اسکیزوفرنیا[6] و اعتیاد می‌شود.

شواهدی مبنی بر نقش دوپامین در تاثیر پاداش‌دهی و ایجاد حس لذت وجود دارد. دسته‌ی اول از این شواهد مبتنی بر تاثیر پاداش‌دهی مواد اعتیادآور هستند. شواهد نشان می‌دهد که پاداش‌ده بودن این مواد و ایجاد حس لذت توسط این مواد به واسطه‌ی زیاد کردن مقدار دوپامین توسط این مواد در مغز است. این امر به عنوان شاهدی بر نقش دوپامین در پاداش‌دهی به شمار می‌رود. در فصل بعد به طور مفصل‌تر این تاثیر مواد اعتیادآور را توضیح می‌دهیم. دسته‌ی دوم از شواهد مبتنی بر تحریک پاداش مغزی[7] هستند. در اینگونه آزمایش‌ها حیوان یاد می‌گیرد که برای دریافت تحریک الکتریکی مغزی، عمل خاصی را انجام دهد. تحریک الکتریکی، اغلب در حوزه‌ی نورون‌های دوپامینی (مانند VTA) انجام می‌گیرد. آزمایش‌ها نشان می‌دهد که حیوان دریافت یک شوک الکتریکی را در این حوزه به پاداش‌های طبیعی[8] مانند غذا ترجیح می‌دهد. این آزمایش‌ها به عنوان شاهدی مبنی بر نقش دوپامین در پاداش بیان شده‌اند.

[1] Substantia nigra pars compacta (SNc)
[2] Striatum
[3] Nucleus accumbens
[4] Prefrontal cortex
[5] Parkinson
[6] Schizophrenia,
[7] Brain stimulation reward (BSR)
[8] Natural rewards



شکل ۲-۱ برگرفته از [۲۷]. نظام دوپامینی در مغز میانی. نورون‌های دوپامینی نشات گرفته از VTA و SNc به بخش‌های مختلفی از مغز مانند NA نگاشته می‌شوند.

با تکیه بر شواهد فوق، تئوری‌های مختلفی برای نقش دوپامین در پاداش‌دهی مغزی ارائه شده‌اند. به عنوان مثال بیان شده است که دوپامین به طور مستقیم مسئولیت تعیین لذت بخش بودن[۱] یک پاداش طبیعی را برعهده دارد [۲۸]. بدین معنا که آزاد شدن دوپامین به هر دلیلی به معنای لذت بخش بودن آن دلیل است. این تئوری کلاسیک دو شاهد پیش را به این صورت توضیح می‌دهد که سوء مصرف مواد به طور مصنوعی دوپامین را زیاد و ایجاد لذت می‌کند. همچنین تحریک الکتریکی مغز باعث آزاد شدن دوپامین در مغز شده و موجب پاداش‌ده بودن این عمل برای حیوان می‌شود. در نتیجه حیوان برای دریافت این پاداش تلاش می‌کند.

نقش دوپامین در پاداش بعدها با جدا کردن مسیرهای[۲] مغزیِ "خواستن"[۳] و "دوست داشتن"[۴] به چالش کشیده شد [۲۹]. این تئوری مبتنی بر مجموعه‌ای از مشاهده‌ها است که نشان می‌دهد گاهی حیوان برای بدست آوردن پاداش طبیعی‌ای که از آن لذت نمی‌برد، تلاش می‌کند (اینکه حیوان لذت می‌برد یا نه، به وسیله تغییرات چهره حیوان در حین مصرف و میزان مصرف اندازه‌گیری شده است). بدین معنا که اگر مسیر "دوست داشتن" مغزی آسیب‌ببیند، حیوان همچنان برای بدست آوردن

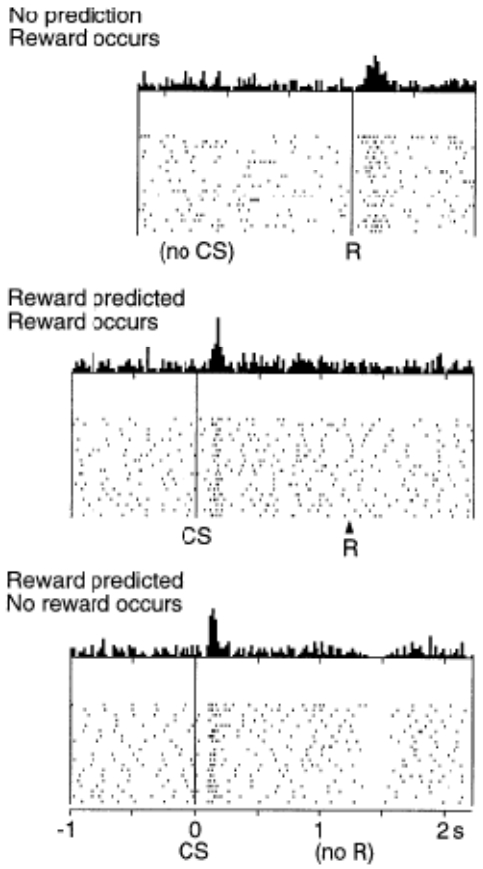
[۱] Hedonic effect
[۲] Pathway
[۳] Wanting
[۴] Liking

پاداش تلاش می‌کند ولی در هنگام مصرف علائم لذت بردن در آن مشاهده می‌شود. ولی درصورتی که مسیر "خواستن" خراب شود، حیوان برای بدست آوردن پاداش تلاش نمی‌کند (مثلا حاضر نیست که مسیر مورد نیاز را برای رسیدن به منبع ساکروز طی کند)؛ این در حالی است که اگر ساکروز به آن داده شود، به میزانی که قبلا مصرف می‌کرد، مصرف می‌کند و علائم لذت را از خود نشان می‌دهد و بنابراین در "دوست داشتن" خللی ایجاد نشده است. این توصیف در کنار این فرض که دوپامین در مسیر "خواستن" قرار دارد و نه "دوست داشتن" نتیجه می‌دهد که دوپامین نقشی در لذت بخش بودن یک مصرف ندارد.

در مقابل هر دو تئوری بیان شده در بخش پیش، شولتز[1] و همکاران در مجموعه‌ای از آزمایش‌ها دیدگاه دیگری ارائه کردند [۳۰] که نقطه‌ی اتصال نقش دوپامین به یادگیری تفاضل زمانی به شمار می‌رود. در آزمایش‌های انجام گرفته، گزارش شده است که بخش عمده‌ای از نورون‌های دوپامینی در VTA و SNc به یک پاداش پیش‌بینی نشده و غیرمنتظره پاسخ نشان می‌دهند. بدان معنا که اگر مثلا مقداری آب میوه بدون سابقه قبلی به میمون داده شود، نورون‌های دوپامینی فعال می‌شوند. همچنین نورون‌های دوپامینی هنگام مشاهده شدن یک محرک که داده شدن یک پاداش در آینده را پیش‌بینی می‌کند، فعال می‌شوند. در مقابل این فعالیت در برابر پاداش پیش‌بینی نشده، هنگامی که دریافت یک پاداش مورد انتظار حیوان باشد، نورون‌های دوپامینی پاسخ نمی‌دهند. از طرفی فعالیت نورون‌های دوپامینی وقتی که محرک مورد انتظار داده نشود، به طور لحظه‌ای دچار توقف می‌شود (شکل ۲-۲).

در اینجا نیاز است که توضیحی در مورد فعالیت نورون‌های دوپامینی ارائه شود. فعالیت نورون‌های دوپامینی که در بالا به آن اشاره شد فعالیت فازیک[2] آنها نام دارد. این فعالیت به فعالیت‌های ضربه‌ای با فرکانسی بالاتر از ۳۰ هرتز اشاره می‌کند. در مقابل این نوع فعالیت، فعالیت یکنواخت[3] نورون‌های دوپامینی به ضربه‌هایی با سطحی ثابت و در فرکانسی حدود ۵ هرتز اشاره می‌کند. از تمایز بین این دو نوع فعالیت در بخش‌های بعدی استفاده می‌کنیم.

[1] Schultz
[2] Phasic activity
[3] Tonic activity



شکل ۲-۲ برگرفته از [۳۰]. فعالیت نورون‌های دوپامینی نقش پیش‌بینی کننده‌ی پاداش را دارند. در شکل بالا، بدون سابقه قبلی، آب میوه به حیوان داده شده است و نورون‌های دوپامینی از خود پاسخ نشان داده‌اند. در شکل وسط، نورون‌های دوپامینی به دادن پاداش از خود فعالیت نشان نداده‌اند ولی نسبت به یک محرک که پیش‌بینی کننده پاداش است، از خود پاسخ نشان داده‌اند. در شکل پایین، در زمان مورد انتظار، پاداش به حیوان داده نشده است و فعالیت نورون-ها به طور لحظه‌ای متوقف شده است.

در بخش بعد به توضیح تئوری‌های بیان شده برای مشاهده‌های فوق می‌پردازیم.

۲-۴-۲ مدل‌سازی

بر اساس مشاهده‌های بیان شده در بخش پیش، می‌توان گفت که نقش فعالیت فازیک نورون‌های دوپامینی، پیش‌بینیِ پاداش است. از آنجایی که هنگامی که اتفاق مورد نظر رخ دهد، نورون‌ها از خود فعالیت نشان نمی‌دهند، می‌توان گفت که اینگونه نورون‌ها حامل سیگنال خطای پیش‌بینی می‌باشند. بدین معنا که هرگاه پاداشی بزرگتر از آنچه مورد انتظار است رخ دهد، نورون‌ها از خود فعالیت نشان

داده، هنگامی که پاداش مورد انتظار داده نشود، فعالیت نورون‌ها به طور لحظه‌ای متوقف می‌شود و در صورت درست بودن پیش‌بینی از خود فعالیت نشان نمی‌دهند. این رفتار نورون‌ها به طور کیفی همانند خطای پیش‌بینی است.

این هماهنگی فعالیت نورون‌ها با سیگنال خطای پیش‌بینی، موجب بهره‌گیری از مدل یادیگری تفاضل زمانی برای توصیف فعالیت فازیک آنها شده است و تاکنون چندین مدل محاسباتی پیشنهاد شده است (برای مرور مراجعه شود به فصل دوم [۳۱]). این مدل‌های محاسباتی را می‌توان به دو دسته تقسیم‌بندی کرد. مدل‌های دسته‌ی اول بر مبنای تابع ارزش (۲-۱) بنا شده‌اند. این بدان معناست که فرض شده پیش‌بینی پاداش در افق زمانیِ متناهی صورت می‌گیرد و بنابراین قادر به توضیح آزمایش-هایی هستند که وظیفه‌ی عامل را به سعی‌های[1] مجزا تقسیم کرده‌اند و در نتیجه مدل‌های این دسته قادر به توضیح پیش‌بینی پاداش در دراز مدت نیستند. به منظور تعمیم این مدل‌ها به افق زمانی نامحدود باید از تابع‌های ارزش کاهشی استفاده کرد که در مدل‌های دسته‌ی دوم به آن پرداخته شده است.

به عنوان نمونه‌ای از مدل‌های دسته اول می‌توان به مدل معرفی شده در [۳۲,۳۳] اشاره کرد. این مدل در واقع همان یادگیری تفاضل زمانی معرفی شده در بخش پیشین است که از بردار وضعیت برای بازنمایی محرک‌ها در قالب وضعیت استفاده می‌کند. از آنجایی که در آزمایش‌های دوپامینی علاوه بر وقوع پاداش، زمان آن نیز توسط حیوان پیش‌بینی می‌شود، در این مدل زمان وقوع یک محرک نیز در بردار وضعیت نگه‌داری می‌شود. این بردار وضعیت را می‌توان به این صورت انگاشت که از زمان وقوع محرک (مانند روشن شدن چراغ) تا موقع دادن پاداش به حیوان، زمان به چندین وضعیت تقسیم‌بندی شده است؛ با گذشت هر واحد زمان، عامل وارد وضعیت بعدی می‌شود. همزمان با یادگیری، ارزش وضعیت دریافت پاداش به وضعیت‌های قبلی رسیده و در نهایت به وضعیت مشاهده محرک می‌رسد. به همین ترتیب مقدار سیگنال خطا ابتدا در وضعیت دریافت پاداش زیاد می‌شود و به مرور به زمان مشاهده‌ی محرک منتقل می‌شود. به این ترتیب، مدل در زمانی مشخص، که منتاظر با وضعیتی مشخص است، انتظار دریافت پاداش را دارد. در پژوهش حاضر چون رفتار زمانی حیوان

[1] Trial

موضوع مطالعه نیست، از همان بردار وضعیت عادی (بدون در نظر گرفتن زمان مشاهده‌ی محرک) استفاده می‌شود و مدل را معادل یک یادگیری تقویتی تفاضل زمانی با تابع ارزش (۲-۱) در نظر می-گیریم. این مدل با این فرض که سیگنال خطا، δ_t، متناظر با فعالیت فازیک نورون‌های دوپامینی است، قادر به توضیح مشاهده‌های بیان شده در بخش پیشین است.

در دسته‌ی دوم از مدل‌ها، مدل معرفی شده در [۳۱] قرار دارد. این مدل بر اساس تابع ارزش میانگین پاداش ارزش استوار است، و به دو منظور بنا شده است: در درجه‌ی اول مدل‌های قبلی را به حالت افق زمانی نامحدود تعمیم می‌دهد، و در درجه‌ی دوم توضیحی برای برخی داده‌های متناقض فعالیت نورون‌های دوپامینی ارائه می‌کند.

به لحاظ عصبی، این مدل بر پررنگ کردن نقش فعالیت یکنواخت نورون‌های دوپامینی بنا شده است. در بخش‌های پیشین اشاره شد که نورون‌های دوپامینی دارای دو نوع فعالیت هستند، فازیک و یکنواخت. در واقع فعالیت نورون‌های دوپامینی به این صورت است که دارای یک فعالیت پیش‌زمینه هستند که همان فعالیت یکنواخت آنها است. در شرایطی خاص (مانند مشاهده یک پاداش پیش‌بینی نشده)، اضافه بر این فعالیت یکنواخت، از آنها فعالیت فازیک نیز مشاهده می‌شود. میزان فعالیت فازیک نورون‌ها بستگی به فعالیت پیش‌زمینه آنها دارد، که اساسِ عصبی این مدل است. در این مدل، بیان می‌شود که فعالیت فازیک نورون‌های دوپامینی نسبت به فعالیت پیش‌زمینه‌ی آنها پدید می‌آید، یا به زبان محاسباتی سیگنال خطا، δ_t، نسبت به یک سطح که همان فعالیت یکنواخت نورون‌ها است اندازه‌گیری می‌شود. به طور غیردقیق مثلا اگر یک قطره‌ی آب میوه فعالیت فازیک نورون‌ها را به اندازه ۱۰ واحد زیاد کند، در صورتی که فعالیت یکنواخت نورون‌ها ۲ واحد افزایش یابد، در صورت دریافت یک قطره‌ی آب میوه فعالیت فازیک نورون‌ها به اندازه‌ی ۸ واحد افزایش می‌یابد. با این بیان، منطقی به نظر می‌رسد که مدل بیان شده مبتنی بر نوعی از یادگیری باشد که میزان سینگال خطا در آن نسبت به سطحی اندازه‌گیری می‌شود. با توجه به رابطه‌ی سیگنال خطا در یادگیری تفاضل زمانی در مدل یادگیریِ متوسط پاداش:

$$\delta_t = r_t - \rho_t + \hat{\mathrm{V}}_{s_{t+1}} - \hat{\mathrm{V}}_{s_t}$$

می‌توان این سطح را همان متوسط پاداش، ρ_t، در نظر گرفت. در این صورت نقش محاسباتی دوپامین یکنواخت، در کد کردن میانگین پاداش خواهد بود. علاوه بر توجیه بیان‌شده در بالا، این مدل قادر به توضیح فعالیت نورون‌های دوپامینی در مقابل تنبیه است که از ذکر آنها در اینجا صرفنظر شده است و خواننده می‌تواند به [۳۴] مراجعه کند.

یافتن زیرساخت‌های عصبی مدل یادگیری تقویتی به سیگنال خطا محدود نمانده است، در پژوهش-های بعدی نحوه‌ی انتخاب عمل (سارسا یا Q-یادگیری)، کد شدن نرخ یادگیری و غیره مورد بررسی قرار گرفته‌اند که در اینجا به ذکر آنها نپرداخته خواننده می‌تواند برای مرور به [۳۵] مراجعه کند. همچنین ارتباط مدل یادگیری تفاضل زمانی با تئوری "خواستن" و "دوست داشتن" در [۳۶] مدل-سازی شده است.

۲-۵ جمع‌بندی

در این فصل در ابتدا به معرفی یادگیری تقویتی پرداختیم. در ادامه پس از مروری بر آزمایش‌های رفتاری شرطی شدن، توانایی مدل یادگیری تفاضل زمانی برای توضیح این مشاهده‌ها مورد ارزیابی قرار گرفت. سپس به تشریح اجمالی زیرساخت‌های عصبی مدل یادگیری تفاضل زمانی پرداخته و نقش محاسباتی فعالیت فازیک و یکنواخت نورون‌های دوپامینی بیان شد. همانطور که در ادامه خواهیم دید، بر مبنای تلقی بیان شده در این فصل از نظام تصمیم‌گیری، می‌توان به توسعه‌ی مدل-های محاسباتی برای اعتیاد پرداخت که موضوع بحث فصل‌های آینده است.

فصل ۳

مروری بر مدل‌سازی عصبی- محاسباتی و دارواییِ اعتیاد

۳-۱ مقدمه

در این فصل به مروری بر مدل‌های عصبی محاسباتی و دارواییِ اعتیاد می‌پردازیم. از آنجا که مدل‌های عصبی-محاسباتی مبنای کار پژوهش حاضر هستند، آنها را به نحو مبسوط توضیح می‌دهیم، در مورد مدل‌های دارویی به شرح کلی آنها و ویژگی‌هایی از آنها که در آینده استفاده می‌شوند، اکتفا می‌کنیم.

۳-۲ مدل‌های مبتنی بر عصب‌شناسیِ محاسباتی

مدل‌های این گروه را می‌توان به دو دسته تقسیم کرد. دسته‌ی اول مدل‌های مبتنی بر یادگیری تقویتی بوده و همگی بر اساس مدل ارائه شده توسط ردیش[۱] در [۳۷] استوار هستند. این مدل مبنای مدل معرفی‌شده در این پایان‌نامه نیز هست. تنها مدل دسته‌ی دوم یک مدل اعتیاد به نیکوتین ارائه-شده توسط گوتکین و همکاران[۲] است [۳۸] و بر اساس رویکرد سیستم‌های دینامیکی ساخته شده است. در این بخش ابتدا به شرح مدل ردیش و بحث در مورد آن پرداخته و سپس مدل اعتیاد به نیکوتین را ارائه می‌کنیم.

---

[۱] Redish
[۲] Gutkin

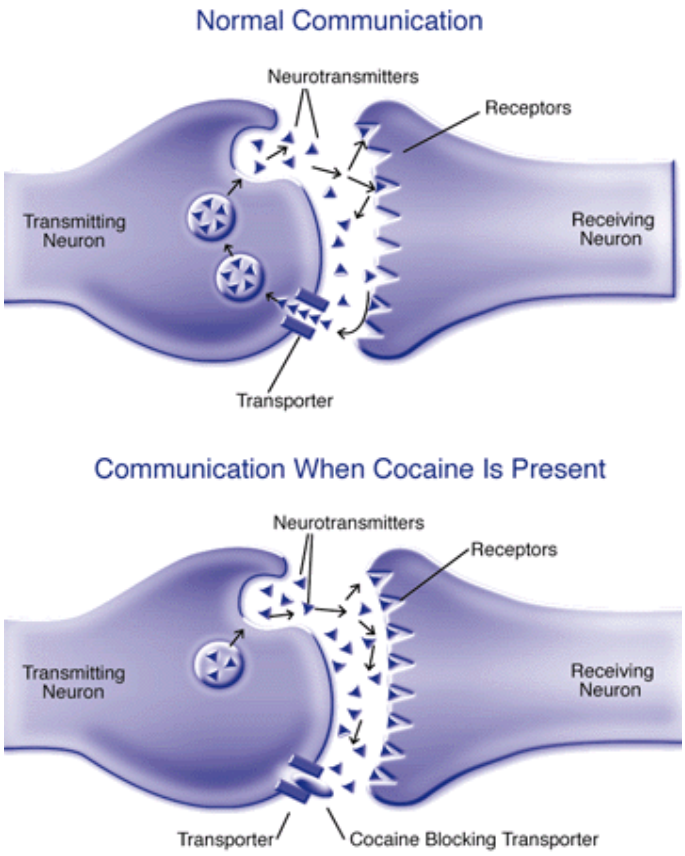
۳-۲-۱ مدل ردیش

عصب‌شناسی

در بخش پیش اشاره شد که مواد اعتیادآور با مداخله در مکانیزم‌های دوپامینی باعث پاداش‌دهی می‌شوند. در واقع تمام مواد اعتیادآور مستقل از مکانیزم عمل و محل عمل، باعث زیاد شدن فعالیت نورون‌های دوپامینی درسیستم مزولیمبیک[1] مغز میانی از جمله VTA شده و در نتیجه باعث زیاد شدن میزان دوپامین در NA می‌شوند [۳۹]. مکانیزم اثر مواد مختلف در تولید این اثر با یکدیگر متفاوت است. در مورد کوکائین نشان داده شده است که اثر پاداشی آن به وسیله‌ی بلوکه کردن منتقل کننده‌های دوپامینی[2] ایجاد می‌شود [۴۰-۴۲] (شکل ۳-۱). موش‌هایی که منتقل کننده‌های دوپامینی در آنها نسبت به کوکائین غیر حساس شده است، از خود شرطی شدن ترجیحِ مکانی[3] نشان نمی‌دهند، که این نشان می‌دهد کوکائین در این موش‌ها اثر پاداش‌دهی نداشته است. زیاد شدن میزان دوپامین همچنین در نمونه‌های انسانی نیز توسط تصویر برداری مغزی PET گزارش شده است [۴۳].

مبتنی بر شواهد بیان شده، مواد اعتیادآور میزان دوپامین را زیاد می‌کنند. این یافته با در نظر گرفتن نقش محاسباتی دوپامین در کد کردن خطای پیش‌بینی نتیجه می‌دهد که مواد اعتیادآور میزان سیگنال خطا را زیاد می‌کنند. برای بررسی این تاثیر بر سیگنال خطا باید به نحو دقیق‌تر به تاثیر کوکائین توجه کرد که در ادامه بیان می‌شود.

[1] Mesolimbic
[2] Dopamine transporter (DAT)
[3] Conditioned place preference



شکل ۳-۱ برگرفته از [۴۲]. انواع مواد اعتیادآور با مکانیزم‌های مختلف باعث زیاد شدن میزان دوپامین می‌شوند. شکل بالا مکانیزم این عمل را در مورد کوکائین نشان می‌دهد. در تصویر بالا چگونگی انتقال پیام از یک نورون به نورون دیگر در شرایط عادی نشان داده شده است. نورون منتقل‌کننده با آزاد کردن ناقل‌های عصبی (نشان داده شده با مثلث) در فضای بین‌سیناپسی باعث جذب ناقل‌های عصبی توسط گیرنده‌های نورون دریافت کننده پیام می‌شود. پس از این مرحله، ناقل‌ها به سمت نورون ارسال‌کننده برگشته و از طریق منتقل کننده‌های عصبی برای استفاده‌های بعدی وارد نورون می‌شوند. شکل پایین، مکانیزم اثر کوکائین را نشان می‌دهد. منتقل کننده‌های عصبی توسط کوکائین بلوک شده، ناقل‌ها قادر به بازگشت به درون نورون فرستنده نیستند و بنابراین مقدار آنها زیاد شده و باعث تحریک بیش از حد نورون دریافت‌کننده می‌شوند.

گرچه مدل ردیش در حالت کلی بیان شده است، با این وجود شواهد ارائه شده بیشتر منطبق بر تاثیرات کوکائین بر نظام دوپامینی است. بررسی دقیق میزان دوپامین در حین خودتزریقی[1] در طول

[1] Self-administration

زمان، سه نوع تاثیر مختلف برای کوکائین مشخص کرده است [۴۴]: فعالیت نورون‌های دوپامینی قبل از پاسخ حیوان (برای دریافت پاداش)، فعالیت نورون‌های دوپامینی بعد از پاسخ (برای دریافت پاداش) و نوع سوم به فعالیت‌های بی‌اختیار لحظه‌ای نورون‌های دوپامینی[۱] اشاره می‌کند. در اینجا قصد ما توضیح این سه نوع پاسخ نیست، لیکن توضیح تاثیر نوع سوم برای فهم مدل ردیش لازم است.

تاثیر نوع سوم به گونه‌ای از فعالیت نورون‌های دوپامینی در اثر مصرف کوکائین اشاره می‌کند که مستقل از مرحله‌ی یادگیری هستند؛ و تنها با غلظت کوکائین در مغز مرتبط هستند. این بدان معنا است که فعالیت این نورون‌ها در طی یادگیری کاهش نیافته (بر خلاف سیگنال خطا در مورد پاداش-های طبیعی که پس از یادگیری کاهش می‌یابد) و مقدار آن تنها تابع میزان غلظت کوکائین در مغز است. از آنجایی که میزان کوکائین گزارش‌کننده‌ی اندازه‌ی سیگنال خطا است، کم نشدن آن در طول یادگیری را می‌توان به معنای صفر نشدن اندازه‌ی سیگنال خطا در طول یادگیری دانست. این بیان، مبنای مدل ردیش است که در بخش بعد بیان می‌شود[۲].

مدل

همان‌طور که در بخش قبل بیان شد، به لحاظ عصبی مدل ردیش بر این مبنا استوار است که سیگنال خطا در طول یادگیریِ ارزش کوکائین به سمت صفر میل نمی‌کند و همواره مثبت است. با این فرض ردیش رابطه‌ی زیر را برای سیگنال خطا پیشنهاد می‌کند:

$$\delta_t^c = \max(\gamma^{d_t}(r_{t+1} + \mathrm{V}_{s_{t+1}}) - \mathrm{Q}_{s_t}^{a_t} + D_{s_t}, D_{s_t}) \qquad (۱-۳)$$

که در واقع همان رابطه (۲-۲۲) است، با این تفاوت که حد پایین سیگنال خطا به D_{s_t} محدود شده است. این بدان معنا است که D_{s_t} بازنمایی کننده تاثیر کوکائین بر میزان دوپامین است. علاوه بر D_{s_t}، در (۳-۱) r_{t+1} نیز در تعیین مقدار سیگنال نقش دارد. در واقع این عبارت برای در نظر گرفتن نقش بالقوه‌ی تاثیرهای کوکائین بر سیگنال خطا از طریق مسیرهای غیردوپامینی است. به این معنا که مقدار سیگنال خطا تنها از تاثیر کوکائین بر میزان دوپامین نشات نمی‌گیرد و از راه‌های دیگری نیز بر

[۱] Spontaneous dopamine transients

[۲] ارتباط نگارنده با دیوید ردیش.

سیگنال خطا تاثیر می‌گذارد. گرچه این نوع تاثیرها تا حدود زیادی ناشناخته است، با این وجود، حضورِ دو عبارت موازیِ پاداش (D_{s_t} و r_t) اشکالی در صحت مدل ایجاد نمی‌کند[1].

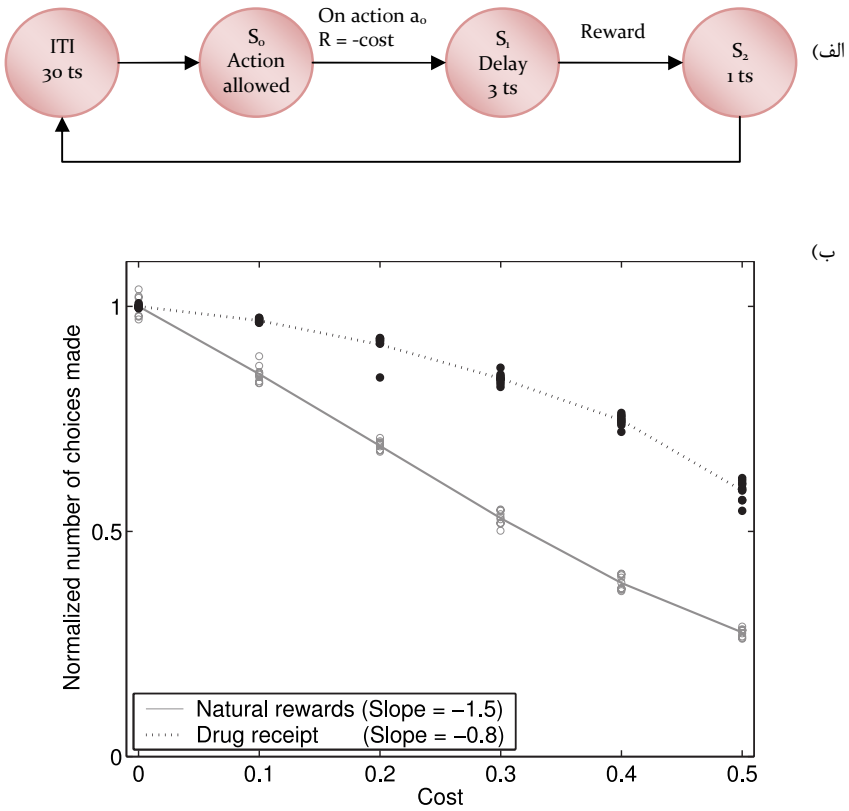
بحث

به صورت شهودی انتظار می‌رفت که زیاد شدن مقدار سیگنال خطا توسط مواد به وسیله‌ی اضافه کردن یک عبارتِ با مقدارِ زیاد به سیگنال خطا صورت گیرد و رابطه‌ای مانند عبارت زیر به عنوان سیگنال خطا در نظر گرفته شود:

$$\delta_t^c = \gamma^{d_t}(r_{t+1} + \mathrm{V}_{s_{t+1}}) - \mathrm{Q}_{s_t}^{a_t} + \phi_{s_t} \qquad (۲-۳)$$

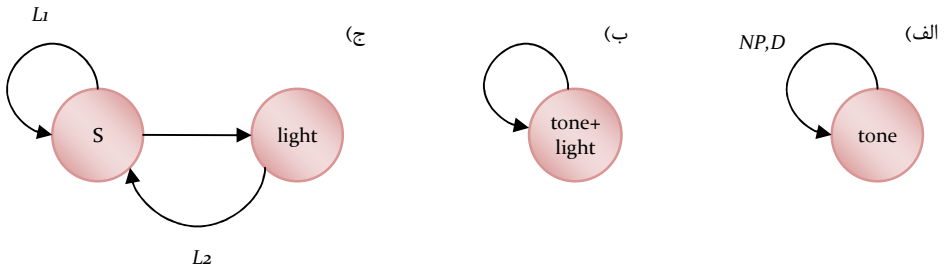
که در عبارت بالا $\phi_{s_t} \gg 0$ و مدل‌کننده‌ی تاثیر مواد است. مشکل تعریف سیگنال خطا به صورت (۲-۳) این است که در ابتدای یادگیری مقدار سیگنال خطا بسیار بزرگ بوده و پس از دفعات محدود مصرف مواد، ارزش وضعیت مصرف مواد زیاد شده به مقدار نهایی خود، که بسیار زیاد است، می‌رسد. بنابراین مدل حتی در مصرف کوتاه‌مدت نیز از خود رفتار غیرحساس به تنبیه نشان می‌دهد. این رفتار با گزارش‌هایی که بیان می‌کنند در مراحل اول اعتیاد، مصرف مواد حساس به تنبیه است، و پس از مصرف درازمدت، تصمیم‌گیری غیرحساس به تنبیه می‌شود، ناسازگار است. به نظر می‌رسد که ردیش به این علت از رابطه‌ی (۲-۳) استفاده نکرده، و سیگنال خطا را به گونه‌ای طراحی کرده که در طول کل یادگیری مقدار ثابت مثبتی داشته و باعث افزایش *تدریجی* ارزش وضعیت مصرف مواد شود. خوشبختانه اینگونه تعریف دارای ایجاب رفتاری مشخصی است که در بخش ۴-۵ توضیح داده می‌شود. به لحاظ عصبی استدلال ردیش برای گذاشتن عملگر بیشینه‌گیری در (۳-۱)، مبتنی بر تاثیر نوع سوم کوکائین است که در بخش قبل بیان شد. شاهد این نوع تاثیر یک سال پس از ارائه‌ی مدل ردیش گزارش شد و به طور کلی شاید نتوان آن را به عنوان یک تاثیر عمومیِ مواد اعتیادآور بر نظام عصبی پذیرفت.

[1] ارتباط نگارنده با سرج آهمد.



شکل ۳-۲ برگرفته از [۳۷]. (الف) محیط شبیه‌سازی به منظور بررسی حساسیت مدل نسبت به هزینه. مدل تا موقعی که عمل a_o را انتخاب نکند در وضعیت s_o باقی می‌ماند. در صورت انتخاب a_o مدل تنبیه دریافت کرده (-cost) و سپس پاداش دریافت می‌کند (مواد و یا پاداش طبیعی). (ب) مقایسه حساسیت مواد و یک پاداش طبیعی به هزینه. نمودار، احتمال انتخاب عمل a_o را در مورد پاداش طبیعی و مواد نشان می‌دهد. همان‌طور که مشاهده می‌شود، انتخاب a_o در مورد مواد حساسیت کمتری به هزینه داشته است.

در مدل ردیش از آنجایی که مقدار سیگنال خطا همواره بزرگتر از صفر است، ارزش پیش‌بینی شده برای مصرف کوکائین در هر بار مصرف حداقل به اندازه αD_{s_t} زیاد می‌شود. این بدان معناست که هرچه یک فرد معتاد بیشتر مواد مصرف می‌کند، مصرف آن برای او ارزشمندتر می‌شود. زیاد شدن ارزش مصرف کوکائین در طول پیشرفت مراحل اعتیاد به معنای کم شدن حساسیت مصرف آن نسبت به هزینه‌ی مصرف است (شکل ۳-۲)؛ بنابراین مدل ردیش توضیحی عصبی برای رفتار غیرحساس به هزینه در معتادان ارائه می‌کند: همزمان با مصرف درازمدت، ارزش وضعیتِ مصرفِ مواد زیاد شده، اثرات سوء مصرف مواد قادر به جبران این ارزش زیاد نیستند؛ بنابراین معتاد برخلاف اثراتِ سوء مواد،



شکل ۳-۳ آزمایش وقوع پدیده بلوک کردن در مورد پاداش کوکائین. در آزمایش انجام شده در [۴۵] دو دسته موش ارتباط وقوع دو محرک را با پاداش کوکائین یاد گرفتند. دسته اول از موش‌ها (موش‌های بلوک شده) در ابتدا ارتباط صدای بوق و پاداش کوکائین را فرا گرفتند (الف). در مرحله‌ی بعد با دو محرک همزمان چراغ روشن و صدای بوق با پاداش کوکائین آزموده شدند (ب). در مرحله‌ی بعد (ج) بررسی شد که آیا ارزش کوکائین به محرک صدای بوق منتقل شده است یا نه. بدین منظور موش‌ها در شرایطی قرار گرفتند که بین فشار دادن دو اهرم انتخاب داشتند (L1, L2). پس از فشار داردن اهرم اول (L1) هیچ محرکی را مشاهده نمی‌کردند ولی در صورت فشار دادن اهرم دوم (L2) محرک چراغ را مشاهده می‌کردند. موش‌های دسته‌ی دوم (موش‌های بلوک-نشده) همانند موش‌های دسته‌ی اول تحت یادگیری قرار گرفته به جز اینکه تجربه‌ی مرحله اول (الف) راطی نکردند. نتیجه‌ی آزمایش آشکار کرد که موش‌های بلوک شده کمتر از موش‌های بلوک نشده در مرحله‌ی سوم، اهرم دوم را انتخاب کردند. این بدان معنی است که در گروه بلوک‌شده، ارزش کوکائین به محرک چراغ سرایت نکرده و محرک بوق، محرک چراغ را بلوک کرده است و بنابراین پدیده‌ی بلوک کردن در مورد کوکائین رخ می‌دهد و پیش‌بینی مدل ردیش در این مورد درست نبوده است.

به مصرف آنها مبادرت می‌کند. با این توصیف، مدل به نحو رضایت‌بخشی توضیحی درباره‌ی مصرف و جستجوی اجباری مواد به عنوان یکی از شاخصه‌های مهم اعتیاد (بیان شده در بخش ۱-۵-۱) در سطح عصبی ارائه می‌کند.

مدل ردیش یک پیش‌بینی رفتاری قابل بررسی ارائه می‌کند. در بخش ۲-۳ پدیده‌ی بلوکه کردن و ارتباط آن با مدل یادگیری تفاضل زمانی را توضیح دادیم. همان‌طور که گفته شد، محرک اول شرطی شده، از شرطی شدن محرک دوم با پاداش جلوگیری می‌کند. این امر به این علت بود که هنگامی که یک محرک کاملا وقوع یک پاداش را پیش‌بینی می‌کند، محرک دیگری که وقوع همان پاداش را پیش‌بینی می‌کند، با پاداش شرطی نمی‌شود. در مورد مواد اعتیادآور از آنجا که بر اساس مدل ردیش

مقدار سیگنال خطا همواره غیرصفر است، هیچ محرکی وقوع یک پاداشِ از نوع مواد را نمی‌تواند کاملا پیش‌بینی کند. بنابراین مدل پیش‌بینی می‌کند که پدیده‌ی بلوکه کردن در مورد کوکائین روی ندهد. این پیش‌بینیِ رفتاری مدل، مورد آزمایش قرار گرفته و شواهد آن را تایید نمی‌کنند. بدین معنا که بلوکه کردن در مورد کوکائین نیز رخ می‌دهد [۴۵] (شکل ۳-۳).

در مدل ردیش از آنجا که مقدار سیگنال خطا همواره بزرگتر از صفر است، مقدار ارزش وضعیت مصرف کوکائین به صورت نامحدود زیاد می‌شود. این ویژگیِ مدل، غیرقابل قبول به نظر می‌رسد چون که اصولا مکانیزم‌های مقابله کننده‌ی طبیعی میزان حداکثر تاثیر کوکائین را محدود می‌کنند. برای رفع این مشکل خود ردیش پیشنهاد می‌کند که به مدل فوق ضریبی با نام *میزان تاثیر گذاری دوپامین* اضافه شود و مقادیر D_{s_t} و r_{t+1} در آن ضرب شده و سپس وارد رابطه سیگنال خطا شوند. مقدار این ضریب در ابتدا شروع مصرف زیاد بوده و به مرور در طول زمان کم می‌شود تا جایی که تاثیر D_{s_t} و r_{t+1} در سیگنال خطا در دراز مدت به صفر می‌رسد. راه حل پیشنهاد شده از زیاد شدن نامحدود مقدار ارزش جلوگیری کرده و مشکل پیاده‌سازیِ مقادیر نامحدود در نظام عصبی را حل می‌کند. با این وجود نقش این پارامتر در سطح الگوریتمی واضح نیست. به این معنا که آیا این متغیر در سطح الگوریتمی در فرایند تصمیم‌گیری نقش دارد یا نه؟ رویکرد ردیش قادر به پاسخ نیست. مشکل دیگری که پیشنهاد ردیش ایجاد می‌کند، در یادگیری پاداش‌های طبیعی است. اگر میزان تاثیر گذاریِ دوپامین به مرور به سمت صفر میل کند، پس باید یادگیری پاداش‌های طبیعی که از طریق دوپامین انجام می‌شود نیز به طور کلی مختل شود، که در واقع این‌گونه نیست.

مشکل دیگر مدل ردیش به کم نشدن تاثیر کوکائین بر نظام دوپامینی در طول دوره‌ی مصرف اشاره می‌کند. بدان معنا که شواهد نشان می‌دهند که پس از مصرف درازمدت کوکائین، این ماده توانایی اولیه‌ی خود را برای زیاد کردن میزان دوپامین از دست می‌دهد [۴۶,۴۷]. این خصلت در مدل ردیش وجود ندارد و کوکائین در تمام مراحل اعتیاد تاثیر پیشین خود را برای زیاد کردن دوپامین حفظ می‌کند.

مشکل دیگر مدل ردیش مربوط به *پدیده‌ی بازگشت*[1] است. بازگشت به این پدیده اشاره می‌کند که حیوان پس از ترکِ رفتارِ پاسخ برای مواد، در شرایط خاصی بدون یادگیری مجدد شروع به پاسخ می-کند. مثلا فرض کنیم حیوان یاد گرفته است که فشار دادن اهرمی با تزریق مواد همراه است. در این شرایط حیوان به طور مداوم اهرم را برای دریافت مواد فشار می‌دهد. حال فرض کنیم در مرحله‌ی بعد حیوان پس از فشار اهرم مواد دریافت نکند. در این صورت حیوان پس مدتی دیگر اهرم را فشار نمی-دهد. این دو رفتار در مورد پاداش‌های طبیعی نیز وجود دارد. اما در مورد مواد، اگر پس از ترک رفتار، مقداری کوکائین به حیوان تزریق شود، تحت استرس قرار گیرد (مثلا با دادن شوک الکتریکی) و یا محرک مربوط شده به مواد مشاهده کند، شروع به فشار دادن اهرم می‌کند [۴۸]. این پدیده بازگشت نام دارد. مدل ردیش پیش‌بینی می‌کند که در صورت عدم داده شدن مواد، ارزش عمل فشار دادن اهرم به سمت صفر میل کرده، کسب مجدد عمل نیاز به یادگیری مجدد دارد. این پیش‌بینی با پدیده-ی بازگشت در تضاد است. در کارهای بعدی، ردیش به منظور برطرف کردن این مشکل از مکانیزم توسعه وضعیت استفاده کرد که در اینجا به توضیح آن نمی‌پردازیم و خواننده می‌تواند به [۴۹] مراجعه کند.

شاید مهمترین ایرادی که به مدل ردیش وارد است، مربوط به تصمیم‌گیری در برابر پاداش‌های طبیعی است. مدل ردیش پیش‌بینی می‌کند که در شرایطی که تصمیم‌گیری در مورد پاداش‌های طبیعی است، نظام تصمیم‌گیری به درستی کار می‌کند. این پیش‌بینی به وضوح برخلاف ویژگی دوم اعتیاد است که در بخش ۱-۵-۱ بیان شد.

۳-۲-۲ مدل گوتکین و همکاران

مدل گوتکین و همکاران یک مدل محاسباتی برای اعتیاد به نیکوتین است که توسط رویکرد سیستم-های دینامیکی توسعه یافته است. به لحاظ عصبی مدل گوتکین برای سیگنال یکنواختِ دوپامین نقش کنترل حافظه را در نظر می‌گیرد. بدان معنا که برای یک یادگیریِ کارا باید سیگنال یکنواخت دوپامین از یک سطح مشخص بالاتر باشد. منظور از یادگیری، برقراری ارتباط بین یک محرک و یک

[1] Reinstatement

عمل است. در مدل گوتکین نقش نیکوتین بر میزان دوپامین در سه افق زمانی مختلف مدل‌سازی شده است. تاثیر نیکوتین بر میزان دوپامین از طریق گیرنده‌های نیکوتین[1] (که آن را با nAChRs نشان می‌دهیم) انجام می‌شود. در مقیاس زمانی اول، نیکوتین به طور غیرمستقیم (از طریق فعال کردن nAChRs) باعث افزایش مقدماتی فعالیت نورون‌های دوپامینی می‌شود. این تاثیر تا چند ثانیه بعد از مصرف دوام پیدا می‌کند. در مقیاس زمانی دوم، مصرف چندباره نیکوتین باعث زیاد شدن پایدار فعالیت نورون‌های دوپامینی می‌شود که برای چند دقیقه باقی می‌ماند. این دو تاثیر دوپامین مانند تاثیر بیان شده در بخش قبل است و موجب انتخاب زیاد عمل مصرف نیکوتین می‌شود. تاثیر نوع سوم در افق زمانی درازمدت به وجود می‌آید. در این نوع تاثیر در اثر فعال شدن زیاد nAChRs، این گیرنده‌ها کم فعالیت شده باعث افت میزان فعالیت یکنواخت نورون‌های دوپامینی می‌شود.

از نظر ریاضی مدل گوتکین بر شبکه عصبی مصنوعی "برنده همه را می‌برد"[2] استوار است. نقش نیکوتین در افق زمانیِ اول و دوم در تقویتِ عملِ انتخاب مواد است. در مدل گوتکین میزان یکنواخت دوپامین در دراز مدت افت می‌کند و موجب ناکارامدی یادگیری می‌شود. این را می‌توان با کمی مصامحه به این معنا در نظر گرفت که در یک مدل یادگیری تقویتی پارامتر نرخ یادگیری بسیار کم شده و مقادیر ارزش‌ها از تجربیات تاثیر نمی‌پذیرد. با این وصف، مدل گوتکین رفتار عدم حساسیت به هزینه را این‌گونه توجیه می‌کند که در مراحل اولیه، فعالیت زیاد نورون‌های دوپامینی باعث گرایش رفتار به سمت انتخاب ماده می‌شود. پس از این گرایش، گرچه هزینه‌های سلامت و اجتماعی و غیره بر معتاد تحمیل می‌شود، لیکن به دلیل ناکارمدی یادگیری، رفتار معتاد تغییر نکرده، همچنان بر رفتار غیربهنیه خود (مصرف مواد) پافشاری می‌کند.

مدل گوتکین فرایند اعتیاد را در سطح نورونی توضیح می‌دهد و از این لحاظ بر مدل ردیش برتری دارد. به علاوه مشکل مقادیر نامحدود ارزش وضعیت‌ها را نیز ندارد. با این وجود این مدل نیز مشکلاتی دارد. در حقیقت مدل گوتکین پس از مصرف درازمدت قادر به ترک رفتار نیست، چه در مورد مواد چه در مورد پاداش طبیعی. بدان معنا که پس از مصرف طولانی نیکوتین و از کار افتادن یادگیری، اگر

[1] Nicotinic acetylcholine receptors (nAChRs)
[2] Winner-take-all

مثلا دیگر رفتار فشار دادن اهرم با دادن نیکوتین همراه نباشد، مدل همچنان عمل فشار دادن اهرم را انتخاب می‌کند. زیرا مدل قادر به یادگیری این نیست که عمل فشار دادن دیگر با دادن مواد همراه نیست. از این جهت مدل قادر به توضیح پدیده‌ی ترک رفتار و به تبع بازگشت نیست.

همچنین، مدل گوتکین پس از مصرف طولانی مواد، کاهش سرعت یادگیری برای پاداش‌های طبیعی پیش‌بینی می‌کند، ولی در صورت یادگیری کاهش انگیزه برای پاداش طبیعی را پیش‌بینی نمی‌کند. بنابراین مدل گوتکین همچون مدل ردیش قادر به توضیح ویژگی دوم اعتیاد نیست.

۳-۳ مدل‌های محاسباتی دارویی اعتیاد

مدل‌های دارویی اعتیاد برخلاف مدل‌های عصبی-محاسباتی بیان شده در بخش قبل، بر قدرت تقویتی مواد[1] تکیه نمی‌کنند. توضیح اینکه پاسخ حیوان برای بدست آوردن پاداش‌های طبیعی بر اثر قدرت تقویت‌کنندگی آنها است. به این معنا که مثلا با زیاد کردن دوز ساکروز در محلول، ارزش وضعیت مصرف پاداش زیاد شده، نرخ پاسخ حیوان بر روی اهرم به طور متناسب زیاد می‌شود. در مورد مواد، شرایط تا حدی متفاوت است. بعد از یک باز تزریق مواد، بلافاصله حیوان دچار سیری[2] شده، انگیزه خود را برای تزریق بعدی از دست می‌دهد. در این دوره‌ی سیری، مواد قدرت تقویت‌کنندگی خود را از دست داده و حیوان برای آن تلاش نمی‌کند. پس از گذشت دوره سیری حیوان مجددا میل به مواد پیدا کرده و برای بدست آوردن آن تلاش می‌کند، مثلا اهرمی را فشار می‌دهد. این ویژگی باعث می‌شود که نرخ پاسخ بر حسب دوز مواد حالت U شکل پیدا کند. بدان معنا که در دوزهای پایین که اثر سیری کم است نرخ پاسخ برای مواد متناسب با دوز آن است ولی در دوزهای بالا با زیاد شدن اثر سیری، این تاثیر معکوس شده، در دوزهای بالاتر نرخ پاسخ کاهش می‌یابد [۵۰].

تمام مدل‌های مطرح شده در بخش قبل از تاثیر سیری مواد صرف نظر کرده و معتاد را در شرایطی در نظر می‌گیرند که فاصله‌ی بین مصارف به اندازه کافی زیاد بوده، در نتیجه مصرف با سیری همراه نشده و مواد اثر تقویت‌کنندگی خود را حفظ می‌کند. بر خلاف آنها، مدل‌های دارویی سعی در مدل‌سازی شرایطی دارند که تزریق‌های متوالی به سرعت قابل انجام است. بنابراین مصرف مواد با سیری همراه

[1] Reinforcing efficacy
[2] Satiety

است. همانطور که به نظر می‌رسد توسعه‌ی این مدل‌ها در این شرایط خاص باعث محدود شدن آنها به توضیح مشاهده‌های آزمایشگاهی می‌شود. از طرفی موضوع مورد توضیح در اینگونه مدل‌ها به طور مستقیم به ویژگی‌های رفتاری بیان‌شده در بخش قبل ۱-۵-۱ مربوط نیست. با این وجود چون از مفاهیم بیان شده در این مدل‌ها در فصل بعد استفاده می‌شود، به توضیح اجمالی آنها و پدیده توصیف‌شده در هر کدام می‌پردازیم.

۳-۳-۱ مدل آهمد و کوب

مدل آهمد و کوب [۵۱] یک مدل محاسباتیِ دارویی است. موضوع مورد توضیح در این مدل مبتنی بر آزمایش‌های نرخ پاسخ موش‌ها در یک برنامه‌ریزی ثابت[1] (که آن را با FR نشان می‌دهیم) برای دریافت مواد است. در یک برنامه‌ریزی FRn (با نسبت n)، حیوان پس از n بار پاسخ (مانند فشار دادن اهرم) تقویت کننده دریافت می‌کند (که در اینجا تقویت کننده مواد است).

مشاهده‌های مبنای این مدل گزارش می‌کنند که نرخ پاسخ موش‌ها برای دوز ثابتی از مواد پس از مصرف طولانی مدت (۶ ساعت در روز) بیشتر از مصرف کوتاه مدت (۱ ساعت در روز) است [۵۲,۵۳]. بدین معنا که اگر موش‌ها برای مدتی (مثلا یک ماه) در یک روز ۶ ساعت دسترسی به مواد داشته باشند، پس از این مدت اهرم را با نرخ بیشتری نسبت به موش‌هایی که در روز ۱ ساعت دسترسی به مواد داشته‌اند فشار می‌دهند. این مشاهده در کنار این تلقی که حیوان هنگامی اقدام به تزریق می‌کند که تاثیر تزریق پیشین از "حد مشخصی" کاهش یافته باشد، این نتیجه را می‌دهد که آن "حد مشخص" پس از مصرف طولانی مدت بالاتر رفته، یکبار تزریق، تاثیر خود را زودتر از دست می‌دهد و از "حد مشخصی" پایین‌تر می‌رود. حد مشخص بیان شده به نام *آستانه‌ی پاداش*[2] شناخته می‌شود و به لحاظ عصبی به آستانه‌ای از پاداش اشاره می‌کند که قادر به فعال کردن نورون‌های دوپامینی است.

پس به طور خلاصه مدل آهمد و کوب در مرحله‌ی اول بیان می‌کند که رفتار خودتزریقی در حیوان به هدف نگه داشتن سطح ماده بالاتر از یک حد مشخص است. در مرحله‌ی بعد بیان می‌کند که با مصرف دراز مدت آن سطح مشخص بالا می‌رود.

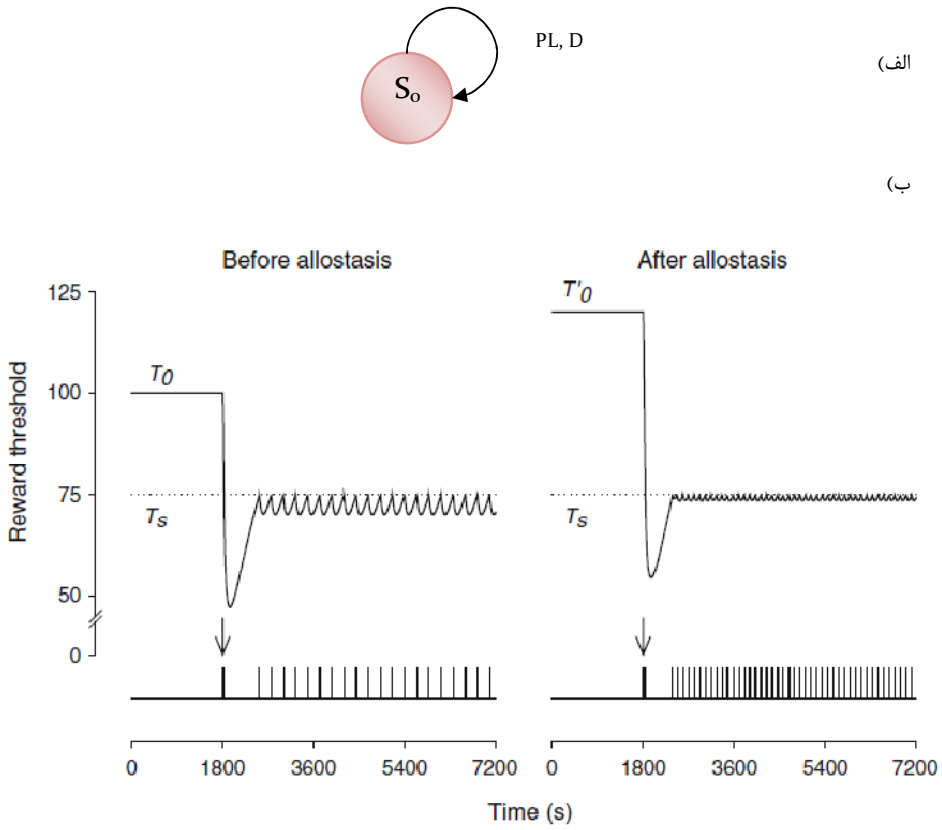
[1] Fixed-ratio schedule (FR)
[2] Reward threshold

از دیدگاه مدل‌سازی، مدل آهمد و کوب در درجه‌ی اول دینامیک کم شدن تاثیر دارو بر بدن را پس از یکبار تزریق مدل می‌کند. در مرحله‌ی بعد چگونگی تغییر آستانه‌ی پاداش را توضیح داده و از طریق شبیه‌سازی نشان می‌دهد که مدل قادر به پیش‌بینی نرخ پاسخ پس از مصرف طولانی‌مدت و کوتاه-مدت مواد است (شکل ۳-۴).

۳-۳-۲ مدل نورمان و تسیبولسکی

این مدل نیز همچون مدل قبلی یک مدل دارویی است [۵۴]. به لحاظ تجربی این مدل بر اساس مجموعه‌ای از آزمایش‌ها استوار است که بیان می‌کنند حیوان، زمانی اقدام به خودتزریقی می‌کند که غلظت کوکائین در بدن در محدوده‌ی مشخصی قرار داشته باشد. این محدوده که *ناحیه‌ی اجبار*[۱] نامیده می‌شود، دارای یک حد پایین و یک حد بالا است. حد پایین آستانه‌ی شروع[۲] و حد بالا آستانه-ی سیری نام دارد. هرگاه غلظت کوکائین در بدن بین این دو حد قرار گیرد، حیوان اقدام به خود-تزریقی مواد می‌کند. آزمایش‌ها نشان می‌دهد که هرگاه به طور مصنوعی غلظت ماده در بدن بین این دو سطح قرار بگیرد، حیوان اقدام به خود تزریقی می‌کند. بنابراین از دیدگاهِ این مدل، خود تزریقی یک فرایند خودکار است که هنگامی که غلظت مواد وارد محدوده‌ی مشخصی شود، صورت می‌گیرد. در این مدل نیز همچون مدل قبلی دینامیک تغییرات غلظت مواد بر بدن مدل‌سازی شده است (بر حسب میزان اولیه، دوز مواد و نرخ تزریق) و سپس بالاتر رفتن سطح آن از آستانه‌ی شروع و پایین‌تر بودن از آستانه‌ی سیری به معنای اقدام به خود تزریقی در نظر گرفته شده است. این مدل در مورد تاثیرهای درازمدت مواد (مانند بالابردن آستانه پاداش) توضیحی ارائه نمی‌کند.

[۱] Compulsion zone
[۲] Priming threshold



شکل ۳-۴ برگرفته از [۵۱]. (الف) برنامه‌ریزی FR1. حیوان پس از یکبار فشار اهرم (PR) مواد دریافت می‌کند (D). (ب) نتایج شبیه‌سازیِ مدل آهمد و کوب. شکل سمت چپ پاسخ‌های مدل را در مصرف کوتاه‌مدت نشان می‌دهد. شکل سمت راست پاسخ‌های مدل را در مصرف درازمدت نشان می‌دهد. پس از مصرف درازمدت و بالارفتن آستانه پاداش، سطح مواد پس از یکبار تزریق سریع‌تر از آستانه‌ی پاداش کمتر می‌شود و بنابراین مدل با سرعت بیشتری اقدام به خودتزریقی می‌کند.

۳-۴ خلاصه

در این فصل به ارائه‌ی مدل‌های مختلف خرد اعتیاد که مبتنی بر عصب‌شناسی محاسباتی و داروشناسی بودند، پرداختیم. در ابتدا مدل ردیش را معرفی کرده و توانایی‌ها و ایرادهای آن را برشمردیم. سپس مدل گوتکین را معرفی کرده و در مورد آن بحث کردیم. در انتها نیز مدل‌های دارویی معرفی شدند. در فصل بعد با ارائه‌ی مجموعه‌ی جدیدی از شواهد عصبی از تاثیر مواد بر نظام عصبی، به ارائه‌ی مدلی جدیدی برای اعتیاد به کوکائین بر مبنای مدل ردیش می‌پردازیم.

فصل ۴

مدل‌سازیِ عصبی- محاسباتی اعتیاد به کوکائین

۴-۱ مقدمه

در فصل گذشته مدل‌های عصبی و دارویی اعتیاد مرور شدند. در این فصل هدف، توسعه‌ی یک مدل عصبی-محاسباتی برای اعتیاد به کوکائین بر مبنای مدل پیشنهادشده توسط ردیش است. در ادامه، ابتدا به ارائه‌ی زیربنای عصبی مدل می‌پردازیم. سپس بر اساس داده‌های عصبی، ساختار مدل را بیان کرده و در انتها رفتار مدل به وسیله شبیه‌سازی ارائه می‌شود.

۴-۲ زیر بنای عصبی

در بخش ۳-۲-۱ بیان شد که مصرف کوکائین باعث زیاد شدن میزان دوپامین می‌شود؛ که با توجه به تناظر میزان دوپامین و سینگال خطا، سیگنال خطا را زیاد می‌کند. این امر موجب زیاد شدن ارزش وضعیت مصرف کوکائین می‌شود. علاوه بر این تاثیر، مصرف درازمدت کوکائین تاثیرهای دیگری نیز بر نظام عصبی می‌گذارد.

در بخش ۳-۳-۱ اشاره شد که شواهد نشان می‌دهد مصرف درازمدت مواد باعث بالا رفتن آستانه‌ی پاداش می‌شود. آستانه‌ی پاداش، آستانه‌ای است که پاداش‌های بالاتر از آن موجب تحریک نورون‌ها می‌شود. همچنین تصویربرداری مغزی نشان می‌دهد که مصرف‌کنندگانِ مزمن کوکائین در برابر

محرک‌های تصویری جنسی، فعالیت مغزی کمتری نسبت به افراد سالم از خود نشان می‌دهند [۵۵]. همچنین کم شدن فعالیتِ مغزی در برابر محرک پول نیز در اثر مصرف درازمدت مواد گزارش شده است [۵۶]. گزارش‌های فوق بیان‌کننده‌ی این واقعیت هستند که مصرف درازمدت مواد اعتیادآور باعث کاهش حساسیت مدارهای مغزی به پاداش می‌شود. این یافته دور از انتظار نیست: پاداش بسیار زیاد مواد اعتیادآور باعث انطباق در نظام عصبی شده و حساسیت آن را کاهش می‌دهد. در نتیجه نظام عصبی تحریک پذیریِ پیشین خود را در برابر پاداش از دست می‌دهد. این نوع تغییر در نظام عصبی فراتر از یادگیری ارزش زیاد برای مصرف مواد است و به تغییری در خود نظام عصبی اشاره می‌کند [۵۷] که در مدل ردیش به آن پرداخته نشده است. در حقیقت این نوع تلقی دارای پیشینه‌ی طولانی در ادبیات اعتیاد است و در تئوری‌هایی مانند "تئوری فرایند مقابله کننده"[۵۸] و جابجا شدن نقطه‌ی تعادل پاداش توسط فعالیت زیاد مکانیزم‌های ضدپاداش [۵۹]، نقش اصلی را برعهده دارد. کاهش حساسیت را می‌توان به صورت کم‌ارزش شدن پاداش یا به عبارتی بالارفتن سطحی که پاداش‌ها نسبت به آن سنجیده می‌شوند، تعبیر کرد. این مطلب را به معنای بالا رفتن آستانه پاداش در نظر می‌گیریم. در ادامه سعی می‌شود با بیان زیرساخت‌های عصبی بالا رفتن آستانه‌ی پاداش، ارتباط این پدیده با مدل یادگیریِ تفاضل زمانی مشخص شود.

به لحاظ عصبی، کم شدن حساسیت بیان‌شده و بالا رفتن سطح پاداش را می‌توان به دو عامل نسبت داد: *کم شدن تعداد دریافت‌کننده‌های دوپامینی* و *بالارفتن غیرعادی سطح فعالیت یکنواخت نورون-های دوپامینی*.

آزمایش‌های تصویر برداری مغزی PET نشان داده‌اند که در معتادانِ با سابقه‌ی طولانی مصرف مواد دریافت‌کننده‌های دوپامینی D2 در استریاتوم (که شامل NA است) تا حد زیادی کاهش می‌یابد [۶۰]. همچنین کاهش دریافت‌کننده‌های دوپامینی در میمون‌هایِ با مصرف مزمن کوکائین نیز گزارش شده است [۶۱]. مهم بودن این یافته‌ها هنگامی آشکار می‌شود که به نقش دریافت‌کننده‌های دوپامینی D2 در پاداش دقت شود. دریافت‌کننده‌های D2 دوپامینی واسطه‌ی پاداش‌ده بودن مواد و پاداش‌های طبیعی هستند [۶۰]. بنابراین طبیعی است که علاوه بر میزان دوپامین آزادشده، تعداد این دریافت‌کننده‌ها نیز در تعیین مقدار سیگنال خطا نقش داشته باشند. در واقع مقدار سیگنال خطا هم

تابع میزان دوپامین آزادشده بوده و هم تابع تعداد دریافت‌کننده‌های دوپامینی است. در نتیجه کم شدن پایدار تعداد گیرنده‌های دوپامینی در اثر مصرف مزمن مواد باعث کم شدن غیرطبیعی میزان سیگنال خطا می‌شود.

از دیدگاهی دیگر بالاتر رفتن آستانه‌ی پاداش در اثر مصرف مزمن مواد، به افزایش سطح فعالیت یکنواخت نورون‌های دوپامینی در NA نسبت داده شده است [۵۱]. گرچه شواهدی مبنی بر زیاد شدن سطح فعالیت‌های یکنواخت نورون‌های دوپامینی پس از مصرف مواد وجود دارد [۶۲,۶۳]، با این وجود در حال حاضر شاهدی مبنی بر ارتباط میان بالا رفتن سطح فعالیت یکنواخت نورون‌های دوپامینی و آستانه‌ی پاداش وجود ندارد[۱].

به طور خلاصه شواهد عصبی نشان می‌دهد که پس از مصرف مزمن مواد، سطحی که پاداش نسبت به آن سنجیده می‌شود به طور غیرعادی بالا می‌رود. این بالارفتن می‌تواند به علت کم شدنِ دریافت-کننده‌های دوپامینی و یا بالا رفتن غیرعادی سطح فعالیت یکنواخت این نورون‌ها باشد. در بخش بعد، از شواهد بیان شده به منظور ارائه‌ی مدلی برای اعتیاد به کوکائین استفاده می‌کنیم.

۴-۳ مدل

مدل ردیش بیان شده در بخش ۳-۲-۱ بر تابع ارزش کاهش نمایی بنا شده است. همانطور که در بخش ۲-۴-۲ بیان شد، استفاده از تابع ارزشِ متوسط پاداش انطباق بیشتری با داده‌های عصبی و رفتاری دارد. بنابراین در ابتدا سعی خواهیم کرد که مدل ردیش را با استفاده از تابع ارزش متوسط پاداش بازنویسی کنیم. همان‌طور که نشان خواهیم داد، بیان مدل ردیش در مدلِ دوپامینیِ مبتنی بر متوسط پاداش، مشکل مقادیر نامحدود ارزش وضعیت‌ها را حل خواهد کرد.

همان‌طور که در بخش ۲-۲-۱ گفته شد، سیگنال خطای تفاضل زمانی در یادگیری متوسط پاداش به صورت است:

$$\delta_t = r_t - \rho_t + \hat{\mathrm{V}}_{s_{t+1}} - \hat{\mathrm{V}}_{s_t} \qquad (۱-۴)$$

[۱] ارتباط نگارنده با سرج آهمد.

از طرفی در مدل ردیش سیگنال خطا در مورد پاداش کوکائین به قرار زیر است:

$$\delta_t^c = \max(\gamma^{d_t}(r_{t+1} + \mathrm{V}_{s_{t+1}}) - \mathrm{Q}_{s_t}^{a_t} + D_{s_t}, D_{s_t})$$

که می‌توان آن را با استفاده از (۴-۱) به صورت زیر نوشت:

$$\delta_t^c = \max(r_{t+1} + \mathrm{V}_{s_{t+1}} - \mathrm{Q}_{s_t}^{a_t} + D_{s_t}, D_{s_t}) - \rho_t \qquad (۴-۲)$$

که در رابطه‌ی فوق مقدار متوسط پاداش، ρ_t، بیرون از عملگر بیشینه‌گیری است. زیرا مربوط به فعالیت فازیک نورون‌های دوپامینی نبوده، کدکننده‌ی فعالیت یکنواخت نورون‌های دوپامینی است. رابطه‌ی بالا به تنهایی برای مدل کردن اثر کوکائین کافی نیست. توضیح اینکه می‌دانیم در مورد پاداش‌های طبیعی متوسط پاداش با استفاده از پاداش‌های دریافت‌شده در هر لحظه، r_t، محاسبه می‌شود:

$$\rho_{t+1} \leftarrow (1-\sigma)\rho_t + \sigma r_t \qquad (۴-۳)$$

لیکن در مورد پاداش کوکائین، پاداش در لحظه‌ی t علاوه بر r_t از مولفه D_{s_t} نیز تشکیل شده است؛ که به طور طبیعی باید در متوسط پاداش نقش داشته باشد. بدین منظور، مقدار پاداش را برحسب سیگنال خطا به صورت زیر می‌نویسیم:

$$r_t = \delta_t - \mathrm{V}_{s_{t+1}} + \mathrm{Q}_{s_t}^{a_t} + \rho_t \qquad (۴-۴)$$

و سپس در قیاس با رابطه‌ی فوق برای مواد خواهیم داشت:

$$r_t^c = \delta_t^c - \mathrm{V}_{s_{t+1}} + \mathrm{Q}_{s_t}^{a_t} + \rho_t \qquad (۴-۵)$$

که در این صورت r_t برای پاداش‌های طبیعی با استفاده از (۴-۳) و در مورد مواد با استفاده از (۴-۵) محاسبه می‌شود و مقدار متوسط پاداش در (۴-۳) با استفاده از آن به‌روز می‌شود.

در واقع تغییرات فوق، تاثیر کوکائین بر مقدار سیگنال خطا را با استفاده از تابع ارزش متوسط پاداش مدل‌سازی می‌کند. در طولِ مصرف، مقدار متوسط پاداش به مرور زمان زیاد می‌شود و باعث خنثی

شدن سیگنال خطا در طول زمان می‌شود و در نتیجه ارزش وضعیت مصرف مواد به طور نامحدود زیاد نمی‌شود. این ادعا در بخش بعدی توسط شبیه‌سازی نشان داده می‌شود.

بازنویسی مدل ردیش بر اساس تابع ارزش متوسط پاداش علاوه بر حل مشکل مقادیر نامتناهی ارزش وضعیت مصرف مواد، این امکان را فراهم می‌کند که بتوان بالا رفتن آستانه‌ی پاداش و کم شدن حساسیت به پاداش‌های طبیعی در مصرف‌کنندگان مزمن مواد را به مدل ردیش اضافه کرد؛ که در ادامه چگونگی آن را توضیح می‌دهیم.

همان‌طور که بیان شد، از دیدگاه کلی، مصرف مزمن مواد اعتیادآور باعث بالا رفتن سطحی می‌شود که پاداش‌ها نسبت به آن سنجیده می‌شوند. مقدار عادی سطحی که پاداش‌ها نسبت به آن سنجیده می‌شوند، بر اساس تئوری بیان‌شدن در بخش ۲-۴-۲، ρ_t است. منحرف شدن مقدار سطح پاداش از این مقدار را می‌توان به وسیله‌ی یک متغیر جدید، مانند κ_t، مدل‌سازی کرد. همزمان با مصرف مواد، مقدار این متغیر زیاد شده و پس از مصرف دراز مدت به حداکثر مقدار خود همگرا می‌شود:

$$\kappa_{t+1} \leftarrow (1-\lambda)\kappa_t + \lambda N \qquad (۶-۴)$$

که در رابطه‌ی فوق N حداکثر انحراف و λ سرعت انحراف را تعیین می‌کنند ($\lambda \ll \sigma$).

با توجه به تغییر بالا در مدل، رابطه‌ی سیگنال خطا به صورت زیر خواهد بود:

$$\delta_t^c = \max(r_{t+1} + \mathrm{V}_{s_{t+1}} - \mathrm{Q}_{s_t}^{a_t} + D_{s_t}, D_{s_t}) - (\rho_t + \kappa_t) \qquad (۷-۴)$$

که در رابطه‌ی فوق $\rho_t + \kappa_t$ سطحی است که پاداش نسبت به آن سنجیده می‌شود. در مراحل اولیه اعتیاد مقدار κ_t نزدیک صفر است و همراه با پیشرفت در مراحل اعتیاد، مقدار آن زیاد شده و باعث بالارفتن آستانه پاداش می‌شود.

همانطور که پیش‌تر بیان شد، کم شدن تعداد گیرنده‌های دوپامینی نقش کاهشی بر مقدار سیگنال خطا دارند. این تاثیر را بدین صورت می‌توان بیان کرد که پس از مصرف مزمن مواد، مقدار سیگنال خطا از مقدار طبیعی آن کمتر خواهد بود. این کم شدن مقدار سیگنال خطا در (۴-۷) دیده می‌شود

[64]. با بالا رفتن مقدار κ_t، مقدار سیگنال نسبت به وضعیت عادی (پیش از مصرف مواد) کمتر می-شود.

از طرفی همان‌طورکه در بخش قبل بیان شد، بالا رفتن آستانه‌ی پاداش به بالا رفتن غیرعادی سطح فعالیت یکنواخت نورون‌های دوپامینی نیز نسبت داده شده است. همین‌طور از بخش ۲-۴-۲ می‌دانیم که ρ_t به وسیله‌ی فعالیت یکنواخت نورون‌های دوپامینی کد می‌شود. بنابراین بالا رفتن غیرعادی سطح این فعالیت‌ها به معنای بالا رفتن غیرعادی اندازه‌ی ρ_t است. این بالا رفتن غیرعادی در (۴-۷) به وسیله‌ی κ_t مدل‌سازی شده است؛ که در واقع با این بیان $\rho_t + \kappa_t$ متناظر با فعالیت یکنواخت نورون‌های دوپامینی است.

به لحاظ محاسباتی می‌توان رابطه‌ی (۴-۷) را به صورت زیر نوشت:

$$\delta_t^c = \max(r_{t+1} + \mathrm{V}_{s_{t+1}} - \mathrm{Q}_{s_t}^{a_t} + [D_{s_t} - \kappa_t], [D_{s_t} - \kappa_t]) - \rho_t \qquad (۴-۸)$$

علت این بازنویسی آن است که بتوان تغییر در تاثیر کوکائین بر میزان دوپامین را در طول زمان مشاهده کرد. همانطور که در (۴-۸) دیده می‌شود مقدار κ_t از D_{s_t} کم شده و حاصل در سیگنال خطا نقش داشته است. این بدان معنا است که در طول مصرف مواد، با زیاد شدن مقدار κ_t، توانایی مواد برای زیاد کردن مقدار سیگنال خطا کم می‌شود و این با گزارش‌های پیشین سازگار است [۴۶,۴۷].

به علت اینکه نظام دوپامینی بین یادگیری پاداش‌های طبیعی و مواد مشترک است، انحراف آستانه‌ی پاداش منجر به تغییر در یادگیری پاداش‌های طبیعی نیز می‌شود. بدان معنا که مقدار سیگنال خطا در مورد پاداش‌های طبیعی به صورت زیر خواهد بود:

$$r_t = \delta_t - \mathrm{V}_{s_{t+1}} + \mathrm{Q}_{s_t}^{a_t} - (\rho_t + \kappa_t) \qquad (۴-۹)$$

در افراد سالم مقدار κ_t برابر صفر بوده، رابطه‌ی (۴-۹) به همان رابطه‌ی (۴-۴) تبدیل خواهد شد. پس از مصرف مزمن مواد، مقدار κ_t رشد کرده و باعث ایجاد اختلال در یادگیری پاداش‌های طبیعی خواهد شده؛ که نتایج رفتاری آن در بخش بعدی شرح داده می‌شود.

جدول ۴-۱ مقادیر پارامترهای شبیه‌سازی

پارامتر	مقدار	معنا
σ	۰.۰۰۵	نرخ به روز کردن متوسط پاداش
D_{s_t}	۱۵	حداقل مدقدار سیگنال خطای کوکائین
α	۰.۲	نرخ یادگیری
μ_N	۵	متوسط اندازه پاداش طبیعی
σ_N	۰.۰۲	انحراف معیار اندازه پاداش طبیعی
μ_{fr}	۲-	متوسط اندازه پاداش بی حرکت بودن
σ_{fr}	۰.۰۲	انحراف معیار پاداش بی حرکت بودن
μ_{sh}	۲۰۰-	متوسط تنبیه شوک
σ_{sh}	۰.۰۲	انحراف معیار تنبیه شوک
μ_c	۲	متوسط پاداش کوکائین
σ_c	۰.۰۲	انحراف معیار پاداش کوکائین
N	۲	حداکثر انحراف متوسط پاداش
λ	۰.۰۰۰۳	نرخ انحراف متوسط پاداش
μ_s	۱	متوسط پاداش کم طبیعی
σ_s	۰.۰۲	انحراف معیار پاداش کم طبیعی
μ_l	۱۵	متوسط پاداش بزرگ طبیعی
ϵ	۰.۱	نرخ اکتشاف و بهره برداری
k	۷	زمان صبر کردن در DDT

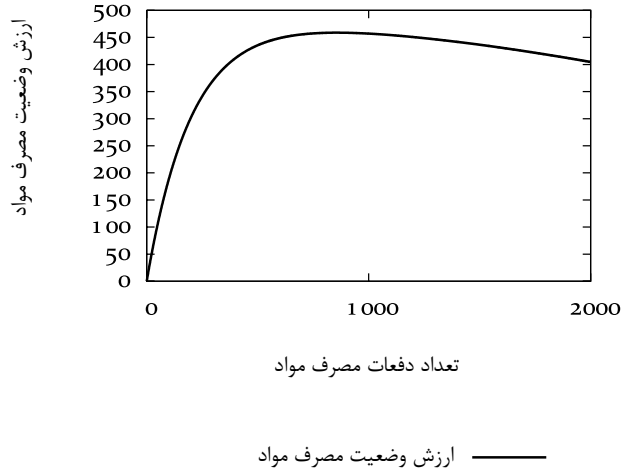
۴-۴ نتایج

۴-۴-۱ جزئیات شبیه‌سازی

تمام پاداش‌ها و تنبیه‌های بیان شده در این بخش دارای توزیع نرمال بوده و مقدار میانگین و واریانس هر یک در جدول ۴-۱ ارائه شده است. پارامترهای توزیع پاداش r_t برای کوکائین با اندیس c نشان داده شده است.

۴-۴-۲ یادگیری ارزش

در شبیه‌سازی اول، یادگیری ارزش مواد در یک برنامه‌ریزی FR1 شبیه‌سازی شد. نتایج شبیه‌سازی در شکل ۴-۱ نشان داده شده است. همانطورکه شکل نشان می‌دهد، ارزش وضعیت مصرف مواد به صورت نامحدود زیاد نمی‌شود. به علاوه نمودار نشان می‌دهد که ارزش وضعیت مصرف مواد پس از رسیدن به یک مقدار بیشینه، افت می‌کند. این افت به دلیل زیاد شدن مقدار κ_t پس از مصرف درازمدت مصرف مواد است. پس از مصرف دراز مدت مواد مقدار ρ_t زیاد شده و به مقدار پاداش دریافت شده میل می‌کند؛ که موجب می‌شود سیگنال خطا به سمت صفر میل کند. با این وجود زیاد



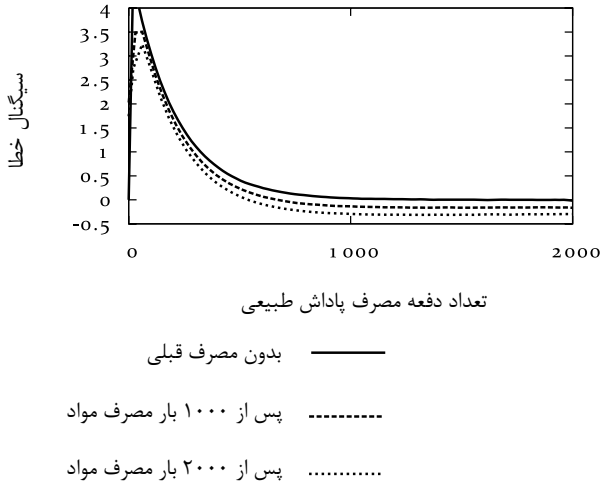
شکل ۴-۱ یادگیری ارزش وضعیت مصرف مواد در یک برنامه‌ریزی FR1. ارزش وضعیت مواد به صورت نامحدود زیاد نمی‌شود. همزمان با مصرف، ارزش آن زیاد شده تا به مقدار بیشینه‌ای می‌رسد و پس از آن به علت بالا رفتن غیرعادیِ آستانه پاداش، افت می‌کند.

شدن مقدار κ_t موجب می‌شود که مقدار سیگنال خطا منفی شده و ارزش تخمینی برای مواد و دیگر تقویت کننده‌ها در طول مصرف افت کند.

شکل ۴-۲ اندازه‌ی سیگنال خطا را درحین یادگیری ارزش یک پاداش طبیعی در یک برنامه‌ریزی FR1 نشان می‌هد. همانطور که دیده می‌شود، با زیاد شدن مدت پیشینه‌ی مصرف، اندازه‌ی سیگنال خطا کم می‌شود و این به علت انحراف آستانه‌ی پاداش با زیاد شدن مقدار κ_t است؛ که در نهایت به کم شدن ارزش تخمین‌زده‌شده برای پاداش طبیعی منجر می‌شود. با این بیان، و با فرض اینکه ارزش وضعیت‌ها و سیگنال خطا متناظر فعالیت بخش‌های مشخصی از مغز هستند [۳۵]، با کم شدن غیرعادی مقدار سیگنال خطا، انتظار می‌رود فعالیت مغز در برابر محرک‌های طبیعی کم شود.

به طور شهودی انتظار می‌رود با کم شدن ارزش پاداش‌های طبیعی در اثر بالارفتن آستانه‌ی پاداش، انگیزه برای دریافت آنها نیز کم شود. به منظور بررسی رفتاری مدل و مشاهده‌ی کاهش انگیزه برای دریافت آن، نیاز است رفتار مدل با آزمایش‌های انجام شده مقایسه شود. به طور کلی می‌توان این آزمایش‌ها را به دو دسته‌ی کلی تقسیم کرد.

در دسته‌ی اول (به عنوان مثال مراجعه شود به [۶۵]) به‌منظور اندازه‌گیریِ انگیزه برای یک پاداش طبیعی (مانند شکر)، از مقایسه میزان انتخابِ آن نسبت به آب استفاده شده است. به این معنا که قبل



شکل ۴-۲ اندازه‌ی سیگنال خطا در حین یادگیری یک پاداش طبیعی در نمونه‌های مدل با پیشینه‌ی متفاوت مصرف مواد. همانطور که نمودار نشان می‌دهد با زیاد شدن مصرف پیشینی مواد، اندازه سیگنال خطا کاهش می‌یابد.

و پس از مصرف درازمدت مواد، نسبت انتخاب عمل مصرف شکر به عمل مصرف آب چگونه تغییر می-کند؛ که گزارش شده است این نسبت کم می‌شود. زیربنای این گونه آزمایش‌ها این است که آب، خود به عنوان یک پاداش نبوده، پس از مصرف دراز مدت مانند بقیه پاداش‌ها دچار افت ارزش نمی‌شود. این فرض در قالب یادگیری تقویتی که بین پاداش‌های مختلف تفاوت نمی‌گذارد در حال حاضر قابل مدل‌سازی نیست. به این معنا که اگر عامل برای مصرف آب عملی را انجام می‌دهد، آب یک تقویت کننده بوده، مانند بقیه‌ی پاداش‌ها پس از مصرف دراز مدت ارزش خود را از دست می‌دهد. برای حل این مشکل می‌توان مسیرهای متفاوتی برای پاداش‌های شیرین مزه مانند شکر و پاداش‌های مانند آب فرض کرد، که مواد تنها مسیر پاداش نوع اول را تخریب می‌کند[۱]. بررسی چگونگی این امر به کارهای آینده موکول می‌شود.

دسته‌ی دوم از این آزمایش‌ها (به عنوان مثال مراجعه شود به [۶۶]) از برنامه‌ریزیِ افزایشی[۲] برای اندازه‌گیری پاداش استفاده می‌کند. در برنامه‌ریزی افزایشی، در صورتی که حیوان n بار پاسخ دهد (مثلا اهرم را فشار دهد)، تقویت‌کننده دریافت می‌کند. در طول آزمایش، n به مرور زیاد شده و حداکثر مقدار n که پس از آن دیگر حیوان اهرم را فشار نمی‌دهد، نقطه‌ی شکست نامیده می‌شود و به

[۱] ارتباط نگارنده با سرج آهمد.

[۲] Progressive-ratio schedule

عنوان شاخصی از انگیزه برای دریافت تقویت‌کننده تلقی می‌شود؛ که نشان داده شده است پس از مصرف درازمدت مواد انگیزه برای پاداش‌های طبیعی کم می‌شود. با کمی ساده‌سازی فرض کنیم که هر بار فشار دادن اهرم، هزینه C_{PL} داشته باشد (هزینه فشار دادن اهرم به علت مصرف انرژی و غیره). در نتیجه n بار فشار دادن اهرم هزینه‌ی nC_{PL} دارد. در این صورت ارزش فشار دادن اهرم برای n بار در افق زمانیِ n برابر خواهد بود با:

$$\mathrm{V}^{PL} = -nC_{PL} + E[R] \qquad (۱۰-۴)$$

با فرض اینکه در صورتی که $\mathrm{V}^{PL} > 0$، حیوان اقدام به فشار دادن اهرم می‌کند، نقطه‌ی شکست (n_{max}) برابر خواهد بود با:

$$n_{max} = \frac{E[R]}{C_{PL}} \qquad (۱۱-۴)$$

حال فرض کنیم پس از مصرف درازمدت، سطحی که پاداش‌ها نسبت به آن سنجیده می‌شود بالا می‌رود (که منجر به کاهش اندازه‌ی پاداش و بالارفتن هزینه فشار دادن می‌شود)، در این صورت نقطه شکست به n'_{max} تغییر خواهد کرد:

$$n'_{max} = \frac{E[R] - \kappa_t}{C_{PL} + \kappa_t} \qquad (۱۲-۴)$$

که نشان می‌دهد با زیاد شدن κ_t، نقطه‌ی شکست کم شده، یا به عبارتی انگیزه‌ی حیوان برای پاداش طبیعی کم می‌شود.

در تحلیل بالا فرض شده که حیوان هنگامی اقدام به فشار دادن اهرم می‌کند که ارزش این‌کار بزرگتر از صفر باشد. این بدان معنا است که فرض شده ارزش گزینه‌ی مقابل فشار دادن (مثلا خوابیدن و یا استراحت کردن) کمتر، مساویِ یا صفر است. همینطور فرض شده است که با بالا رفتن آستانه‌ی پاداش، ارزش گزینه‌ی مقابل فشار دادن اهرم، تغییر نمی‌کند. این فرض، مشابه فرض بیان‌شده در دسته‌ی اول از آزمایش‌ها در مورد تغییر نکردن ارزش آب پس از مصرف درازمدت است. بررسی صحت این فرض نیاز به آزمایش‌های تجربی بیشتری دارد؛ گرچه بدون این فرض نیز کم شدن نقطه‌ی

شکست توسط مدل پیش‌بینی می‌شود. به علاوه فرض شده است که بالارفتن آستانه‌ی پاداش باعث زیاد شدن هزینه‌ی فشار دادن اهرم می‌شود. به این معنا که تنبیه‌ها نیز همچون پاداش‌ها نسبت به سطح پاداش سنجیده می‌شوند. بررسی صحتِ این فرض نیز در حد دانشِ نگارنده در حال حاضر قابل پاسخ نیست.

۳-۴-۴ جستجو و مصرف اجباری مواد

همان‌طور که در بخش‌های قبلی بیان شد، جستجو و مصرف اجباری مواد به این معنا است که معتاد بر خلاف عواقب سوء آینده، به جستجو و مصرف اجباری مواد می‌پردازد. برای بررسی این پدیده تاکنون مدل‌های حیوانی مختلفی طراحی شده‌اند [۶۷-۷۰]. به عنوان مثال در مدل حیوانی بیان شده در [۷۰] حساسیت موش‌ها به شوک الکتریکی قبل و بعد از مصرف درازمدت کوکائین بررسی شده است. در آزمایش بیان‌شده دو اهرم وجود دارد. فشار دادن اهرم اول که اهرم جستجو نام دارد منجر به فعال شدن اهرم دوم می‌شود. با فشار دادن اهرم دوم، موش مواد دریافت کرده و چرخه تکرار می‌شود. در مرحله‌ی بعد که مرحله‌ی آزمایش است، موش در شرایطی قرار می‌گیرد که به طور اتفاقی به او شوک داده می‌شود. در این شرایط موش دو انتخاب دارد. انتخاب اول بی‌حرکت شدن و جمع شدن[1] است که در این صورت تاثیر شوک کمتر خواهد شد، آن را می‌توان به عنوان یک تنبیه کوچک تلقی کرد. انتخاب دوم، فشار دادن اهرم جستجو است. در این صورت چون موش بی‌حرکت نشده است، شک به عنوان یک تنبیه بسیار بزرگ خواهد بود. از انتخاب موش بین این دو عمل می‌توان به اجباری بودن و یا نبودن پاداش‌جویی پی‌برد. در صورتی که موش، فشار دادن اهرم را انتخاب کند، نشان دهنده‌ی این است که موش برای بدست آوردن مواد به تنبیه حساس نبوده و از خود رفتار جستجوی اجباری نشان داده است. در صورتی که موش بی‌حرکت شود و فشار دادن اهرم را انتخاب نکند، این بدان معناست که جستجوی مواد به تنبیه حساس بوده و رفتار پاداش جویی اجباری نشده است. آزمایش انجام شده نشان داد که موش‌ها پس از مصرف درازمدت مواد فشار دادن اهرم را انتخاب می‌کنند، ولی پس از مصرف کوتاه‌مدت مواد به جای فشار دادن اهرم جمع می‌شوند. این رفتار در مورد

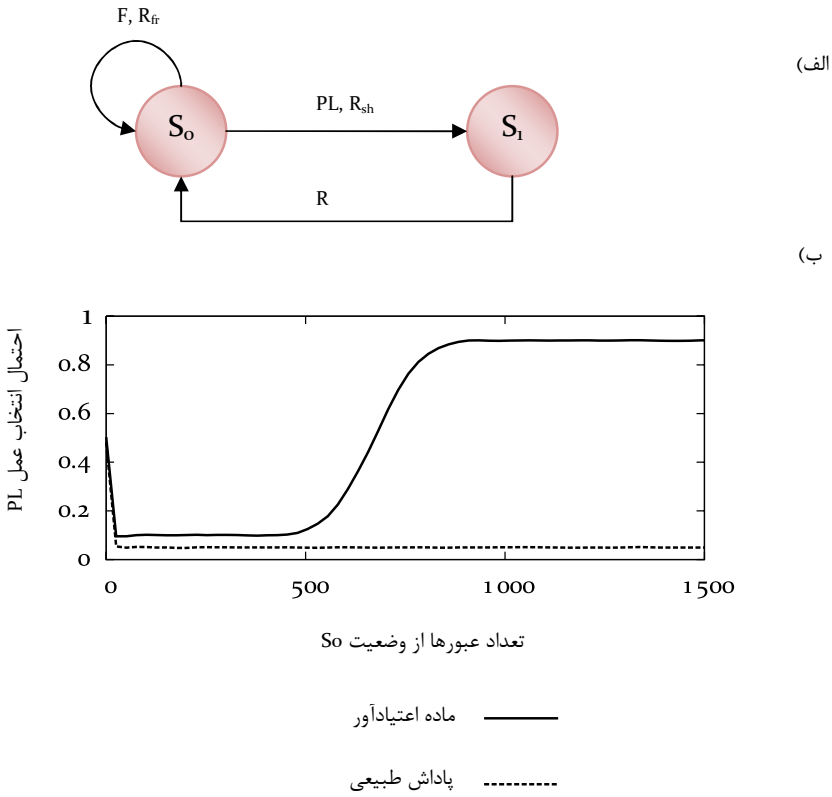
[1] Freezing

ساکروز به عنوان یک پاداش طبیعی مشاهده نشد، پس از مصرف کوتاه‌مدت و درازمدت، موش بی-حرکت‌شدن را انتخاب می‌کرده است.

در اینجا ما مدل را در شرایط مشابهی شبیه‌سازی می‌کنیم (شکل ۴-۳). همان‌طور که شکل نشان می‌دهد پس از مصرف کوتاه‌مدت و درازمدت پاداش طبیعی، مدل رفتار آسیب‌گریزانه‌ی خود را حفظ کرده، عمل بی‌حرکت شدن را انتخاب کرده و اهرم را فشار نداده است. در مورد پاداش مواد، مدل در مراحل اولیه‌ی مصرف عمل فشاردادن اهرم را انتخاب نکرده است، با این وجود پس از مصرف درازمدت مدل عمل فشار دادن اهرم را انتخاب کرده است.

علت عدم حساسیت انتخاب مدل در درازمدت به تنبیه مربوط به زیادشدن پاداش مواد است. این نوع زیادشدن بر اثر تاثیر ماده‌ی اعتیادآور بر میزان دوپامین ناشی می‌شود. چگونگی زیادشدن ارزش وضعیت‌ها در مدل پیشنهادشده در این فصل با مدل ردیش متفاوت است. در مدل ردیش فرض مقدار زیاد برای تاثیر مواد غیرضروری است. بدین‌معنا که اگر مقدار کمی هم برای D_{s_t} فرض شود، پس از مصرف به اندازه کافی طولانی، وضعیت مصرف مواد ارزش زیادی کسب کرده و نسبت به عواقب بد ناشی از آن غیرحساس می‌شود. در مدل پیشنهادشده در این فصل، سینگال خطا به مرور زمان به سمت صفر میل می‌کند. در مدل ردیش این خاصیت وجود ندارد و سیگنال خطا به سمت صفر نمی-کند. در آن مدل اگر سعی شود که سیگنال خطا قابل خنثی شدن طراحی شود (به سمت صفر میل کند)، مثلا با حذف کردن عملگر بیشینه‌گیری، آنگاه مقدار ارزش وضعیت مصرف مواد پس از تعداد محدودی مصرف به سمت مقدار حداکثر خود میل می‌کند. بنابراین در مصرف کوتاه‌مدت نیز از مدل عدم حساسیت به مواد دیده می‌شود. در مدل حاضر مقدار سیگنال خطا قابل خنثی شدن است، لیکن این اتفاق پس از مصرف طولانی روی می‌دهد. در واقع زیاد شدن مقدار ارزش وضعیت هنگامی متوقف می‌شود که داشته باشیم $\delta_t < 0$ که این زمانی تحقق می‌یابد که داشته باشیم $\rho_t + \kappa_t > D_{s_t}$. با توجه به رشد کند ρ_t و بسیار کند κ_t، رسیدن این دو مقدار به مقداری بیشتر از D_{s_t} در مصرف کوتاه-مدت اتفاق نمی‌افتد و این امر توضیح دهنده‌ی رشد ارزش وضعیت مصرف مواد پس از مصرف کوتاه-مدت است. بنابراین رشد ارزش وضعیت مصرف مواد تا درازمدت ادامه پیدا می‌کند و این باعث می‌شود که حساسیت انتخاب مواد به تنبیه به مرور زمان کم شود. در مدل پیشنهاد شده، مقدار نهایی ارزش

وضعیت مصرف مواد وابسته به مقدار D_{s_t} است و فرض یک مقدار زیاد برای D_{s_t} ضروری است. بدین معنی که تنها در صورت زیاد بودن مقدار این عبارت است که در درازمدت حساسیت مصرف مواد به تنبیه به مقدار قابل‌توجهی کم می‌شود. در واقع در مدل پیشنهادشده، زیاد شدن ارزش پاداش‌های طبیعی نیز تا درازمدت ادامه پیدا می‌کند. با این وجود، تفاوت ارزش این پاداش‌ها در درازمدت و کوتاه‌مدت به اندازه‌ای نیست که منجر به کم شدن قابل‌توجه حساسیت مصرف به تنبیه شود. این امر توجیه کننده‌ی رفتار مشاهده‌شده در شکل ۴-۳ در مورد پاداش‌های طبیعی است.



شکل ۴-۳ شبیه‌سازی جستجوی اجباری مواد. (الف) محیط شبیه‌سازی. مدل دو انتخاب دارد: بی‌حرکت شدن (F) و یا فشاردادن اهرم برای دریافت پاداش (PR). در انتخاب F، مدل نه پاداش دریافت می‌کند و نه تنبیه بزرگ شوک و تنها یک تنبیه خیلی کم R_{fr} دریافت می‌کند. در صورتی که مدل PL را انتخاب کند، در ابتدا یک تنبیه بزرگ R_{sh} را دریافت کرده و سپس پاداش R را دریافت می‌کند. (ب) نتایج شبیه‌سازی. محیط فوق برای دو پاداش مواد و پاداش طبیعی (R_N) به طور جداگانه شبیه‌سازی شده است. همان‌طور که نمودار نشان می‌دهد در مورد پاداش طبیعی در تمام طول مصرف، عامل با احتمال کمی PL را انتخاب می‌کند. در صورتی که در مورد پاداش ماده، عامل در کوتاه مدت عمل F را انتخاب کرده و پس از تعداد زیادی مصرف عمل PL را انتخاب کرده است.

۴-۴-۴ تکانشگری

تکانشگری به گونه‌ای از رفتار اطلاق می‌شود که شاخصه آن "عمل‌هایی است که به درستی ادراک نشده‌اند، قبل از پخته شدن ابراز می‌شوند، همراه مخاطره بوده، برای موقعیتی که در آن ابراز می‌شوند مناسب نیستند و اغلب عواقب ناگوار به همراه دارند" [۷۱]. تکانشگری یک پدیده چندوجهی است، با

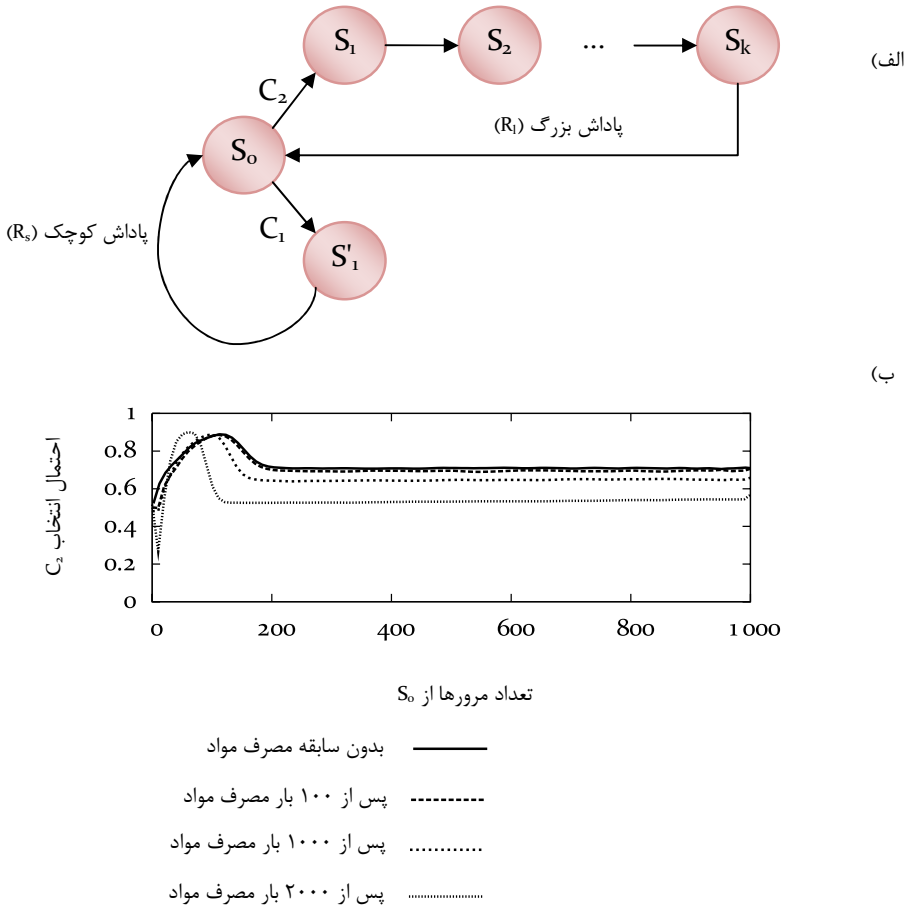
این وجود دو جنبه‌ی از آن با اعتیاد ارتباط نزدیکی دارد: انتخاب تکانشگر[1] و عدم توانایی در انجام ندادن یک عمل[2].

انتخاب تکانشگر به تصمیمی اشاره می‌کند که در آن یک پاداش زودهنگامِ به مقدار کم به یک پاداشِ بزرگِ دیرهنگام ترجیح داده می‌شود. عدم توانایی در انجام ندادن یک عمل به این اشاره می‌کند که در موقعیتی که انجام یک عمل غیربهینه است (مثلا در صورت انجام پاداش دریافت نمی‌کند) و عواقب سوء در پیش دارد، فرد قادر به جلوگیری از انجام عمل نیست.

جهت رابطه‌ی علّی بین اعتیاد و تکانشگری هنوز به طور کامل مشخص نیست. بدین معنا که تکانشگری باعث اعتیاد می‌شود، یا تکانشگری از سوء مصرف مواد ناشی می‌شود؟ برای هر دو شواهدی بیان شده است. به عنوان مثال بیان شده است که مصرف دراز مدت مواد باعث افزایش تکانشگری می‌شود [۷۲]. به علاوه در مدل‌های حیوانی شواهدی مبنی بر افزایش میزان انتخاب تکانشگر وجود دارد [۷۲-۷۴]. از طرف دیگر گزارش‌هایی مبنی بر مقدم بودن خصلت تکانشگری به کسب رفتار معتادانه وجود دارد [۷۵]. در ادامه به شبیه‌سازی شواهد اول، اینکه مصرف مواد موجب تکانشگری می‌شود، می‌پردازیم.

تکانشگری در انتخاب، اغلب به وسیله‌ی آزمون تنزیل تاخیری[3] (که آن را با DDT نشان می‌دهیم) اندازه‌گیری می‌شود. در این آزمون، عامل بین دو گزینه با نتایج مختلف باید انتخاب کند (شکل ۴-۴الف). انتخاب اول (C_1) بلافاصله با یک پاداش کم (R_S) همراه است. در مقابل با انتخاب گزینه‌ی دوم (C_2) عامل باید مدت زمانی صبر کند و سپس به آن یک پاداش بزرگ (R_I) داده می‌شود. تاخیر پس از انتخاب گزینه‌ی دوم به این صورت مدل شده است که عامل باید k وضعیت طی کند، در هر وضعیت یک واحد زمانی، که روی هم می‌شود k واحد زمانی و سپس پاداش R_I دریافت می‌کند. در صورت انتخاب گزینه اول عامل پس از طی یک واحد زمانی پاداش دریافت می‌کند.

[1] Impulsive choice
[2] Impaired inhibition
[3] Delayed discounting task (DDT)



شکل ۴-۴ DDT و شبیه‌سازی رفتار مدل در مراحل مختلف اعتیاد. (الف) DDT. عامل دو انتخاب دارد. می‌تواند گزینه C_1 که در این صورت پس از یک واحد زمانی پاداش کم R_s را دریافت می‌کند. با انتخاب گزینه دیگر، C_2، عامل پس از k واحد زمانی تاخیر پاداش بزرگ R_l را دریافت می‌کند. (ب) رفتار مدل در DDT. با زیاد شدن پیشینه‌ی مصرف مواد، مدل به سمت پاداشِ زودهنگام گرایش پیدا می‌کند.

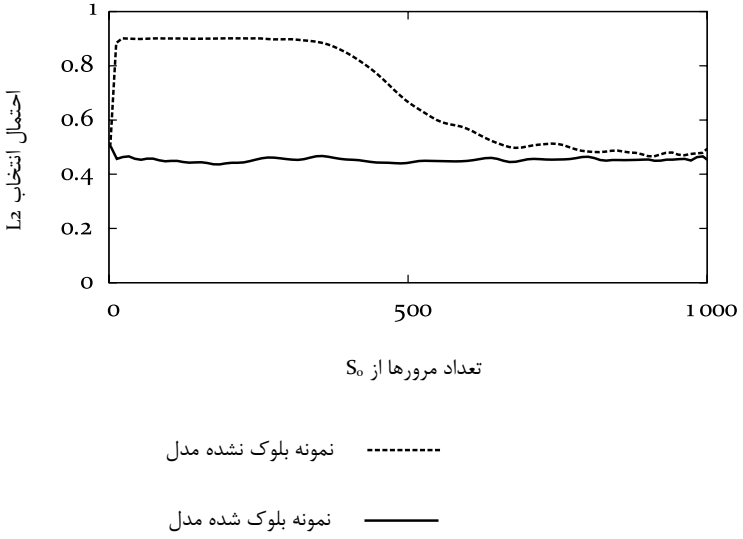
شبیه‌سازی با استفاده از نسخه‌ی یادگیری متوسط پاداش در محیط‌های شبه‌مارکوف صورت گرفت [۷۶]. نتایج حاصل از شبیه‌سازی مدل پس از مصرف پیشین مواد در شکل ۴-۴ب نشان داده شده است. همانطور که دیده می‌شود با زیاد شدن مدت مصرف پیشین مواد، مدل به سمت گزینه با تاخیر کمتر و پاداش کوچکتر گرایش پیدا کرده است. یا به عبارت دیگر رفتار تکانشگری در آن زیاد شده است. علت این امر آن است که در مدل تابع ارزش متوسط پاداش، هزینه‌ی صبر کردن به اندازه‌ی یک واحد زمان توسط ρ_t کد می‌شود. مقادیر زیاد ρ_t به معنای این است که مقدار متوسط پاداش زیاد

است، صبر کردن هزینه‌ی زیادی دارد و بنابراین انتخاب‌های مدل به سمت گزینه‌هایی با پاداش‌دهی سریع هدایت می‌شود. بر عکس مقادیر کم ρ_t به معنای کم بودن نرخ پاداش دریافتی در طول زمان است و بنابراین انتخاب‌های مدل را به سمت گزینه‌هایی با پاداش‌دهی دیرهنگام ولی بزرگ هدایت می‌کند. الگوریتم یادگیری، مدل را سمت انتخاب‌هایی هدایت می‌کند که در کل مقدار متوسط پاداش بیشینه شود. حال فرض کنیم که در اثر مصرف مواد، مقدار موثر ρ_t در سیگنال خطا زیاد شده و به اندازه κ_t افزایش باید $(\rho_t + \kappa_t)$. این بدان معنا است که هزینه‌ی صبر کردن از مقدار واقعی آن بیشتر شده و مدل به سمت انتخاب‌های زودپاداش‌ده گرایش پیدا می‌کند. این دلیل رفتار مشاهده‌شده در شبیه‌سازی است که با داده‌های مبنی بر افزایش رفتار تکانشگرایانه با مصرف مواد مطابقت دارد. بنابراین از این جهت مدل توضیحی برای رفتاری تکانشگرایانه معتادان ارائه می‌کند.

به لحاظ عصبی وجود تکانشگری به کم بودن دریافت‌کننده‌های D2 دوپامینی در NA مربوط شده است [۷۷]. بر اساس مدل پیشنهاد شده، کمبود این دریافت‌کننده‌ها موجب کم شدن سیگنال خطا شده، منجر به تکانشگری می‌شود و از این جهت مدل با این گزارش‌ها سازگار است. با این وجود آزمایش [۷۷] نشان می‌دهد که کم‌بودن دریافت‌کننده‌ها پیش از معتاد شدن وجود دارد و زمینه‌ساز آن است. از این جهت با یافته‌های قبلی مبنی بر کم شدن دریافت‌کننده‌ها مطابقت ندارد و بررسی چگونگی جهت تاثیرگذاری این دو بر یکدیگر نیاز به آزمایش‌های بیشتر دارد.

۴-۴-۵ پدیده بلوکه کردن

همان‌طور که در فصل قبل بیان شده، در مدل ردیش چون مقدار سیگنال خطا در طول مصرف کوکائین به سمت صفر میل نمی‌کند، نمی‌تواند پدیده‌ی بلوکه کردن در مورد کوکائین را توضیح دهد. در مدل پیشنهادشده در این فصل مقدار سیگنال خطا به سمت صفر میل می‌کند و بنابراین طبیعی است که بتواند پدیده‌ی بلوکه کردن را توضیح دهد. برای شبیه‌سازی بلوکه کردن از تقریب خطی ارزش وضعیت بیان شده در بخش ۲-۳-۲ به همراه سیگنال خطای تعریف شده در (۴-۷) استفاده شد. محیط شبیه‌سازی همانند محیط آزمایش شده در [۴۵]، توضیح داده شده در شکل ۳-۳، بوده‌است. بدین منظور دو نمونه از مدل هرکدام متناظر با یک گروه از موش‌ها، شبیه‌سازی شد. بدان معنا



شکل ۴-۵ شبیه‌سازی پدیده‌ی بلوک کردن. در نمونه بلوک شده، پاداش کوکائین با محرک چراغ مربوط نشده و به همین علت مدل با احتمال کمتری عمل L2 را انتخاب کرده است. عمل L2 منجر به رفتن مدل به وضعیت چراغ روشن می‌شده است. در نمونه بلوک نشده، ارزش کوکائین به محرک چراغ منتقل شده و بنابراین مدل عمل L2 را انتخاب کرده است. این نتایج نشان می‌دهد که پدیده بلوک کردن در مورد پاداش کوکائین نیز روی می‌دهد.

که در مورد نمونه‌ی بلوکه‌شده از مدل، در ابتدا محرک بوق (شکل ۳-۳الف) و سپس محرک‌های چراغ و بوق با پاداش مواد مربوط شدند (شکل ۳-۳ب). در نمونه بلوکه‌نشده مدل، تنها محرک چراغ و بوق با پاداش مواد همراه شد. در مرحله‌ی آخر در مورد هر یک از مدل‌ها بررسی شد که آیا محرک چراغ با پاداش مواد مربوط شده است یا خیر. برای این منظور مدل در شرایطی شبیه‌سازی شد که انتخاب L2 منجر به رفتن مدل به وضعیت چراغ می‌شود و با انتخاب L1 مدل بدون دریافت پاداش به وضعیت قبلی خود باز می‌گشت (شکل ۳-۳ج). نتایج حاصل از شبیه‌سازی در شکل ۴-۵ نشان داده شده است. همانطور که دیده می‌شود، نمونه‌ی بلوکه شده L2 را با احتمال کمی انتخاب کرده است؛ این حکایت از سازگاری مدل با وجود پدیده‌ی بلوکه کردن در مورد کوکائین دارد.

همان‌طورکه در بخش قبل بیان شد، مقدار سیگنال خطا، روند کاهشی طی کرده و پس از مصرف طولانی‌مدت به سمت صفر میل می‌کند. از طرفی می‌دانیم که وقوع پدیده بلوکه کردن وابسته به صفر شدن سیگنال خطا است. بنابراین در مدل پیشنهادشده، پدیده‌ی بلوکه کردن هنگامی اتفاق می‌افتد که یادگیری پدیده‌ی بلوکه‌کننده (بوق) به مدت طولانی انجام شده باشد. شبیه‌سازی به این صورت

انجام شده است که یادگیری با محرک بوق به تعداد دفعات زیاد انجام شده است. در صورتی که این یادگیری (یادگیری محرک بلوکه‌کننده) تا صفر شدن مقدار سیگنال خطا انجام نگیرد، مقدار ارزش کوکائین به صورت جزئی به پدیده‌ی بلوکه‌شده (چراغ) منتقل می‌شود. در این صورت دیگر احتمال انتخاب عمل L2 در گروه بلوکه شده صفر نخواهد بود. ولی با این وجود به علت کم شدن سیگنال خطا در طول یادگیری، پس از یادگیری محرک بلوکه‌کننده، مقدار سیگنال خطا کم شده و در نتیجه ارزش مربوط‌شده با محرکِ چراغ در گروه بلوکه‌شده کمتر از بلوکه‌نشده خواهد بود. در نتیجه باز هم نمونه‌ی بلوکه شده‌ی مدل، با احتمال کمتری نسبت به نمونه‌ی بلوکه نشده عمل L2 را انتخاب خواهد کرد. در واقع در آزمایش گزارش‌شده در [۴۵]، نیز همین پدیده بیان شده است که گروه بلوکه‌شده از موش‌ها کمتر از گروه بلوکه‌نشده به سمت محرک چراغ گرایش دارند.

شبیه‌سازی پدیده‌ی بلوکه کردن علاوه بر نیاز به بازنمایی محرک چندگانه در بردار وضعیت، دارای پیچیدگی بیشتری نیز هست. این پیچیدگی از چندمرحله‌ای بودن این آزمون نشات می‌گیرد. بدین معنا که پس از مرحله‌ی اول (یادگیری محرک بلوکه کننده)، باید علاوه بر ارزش وضعیت‌ها، مقدار متوسط پاداش نیز به مرحله‌ی بعد منتقل شود. زیرا که ارزش وضعیت‌ها وابسته به متوسط مقدار پاداش هستند. برای این منظور در شبیه‌سازی انجام شده، به طور ساده در ابتدای مرحله‌ی دوم مقدار متوسط پاداش از مرحله‌ی اول منتقل شد. انجام این عمل به منظور شبیه‌سازی پدیده‌ی بلوکه کردن در مورد پاداش‌های طبیعی نیز لازم است.

۴-۵ پیش‌بینی

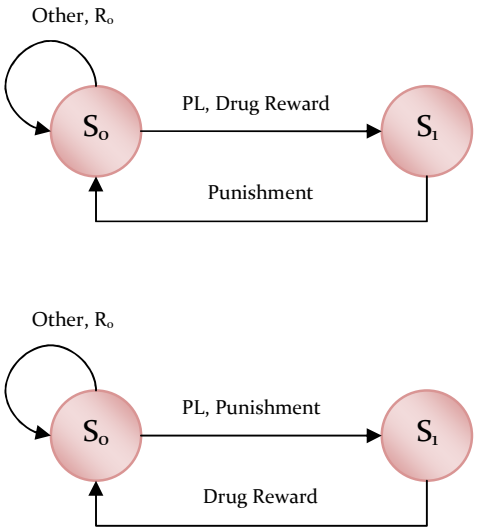
در رابطه‌ی یادگیری پاداش مواد در مدل ردیش (۴-۲) و در مدل پیشنهادی (۴-۷)، وجود عملگر بیشینه‌گیری باعث حساس بودن مدل به ترتیب زمانی پاداش مواد و تنبیه می‌شود. بدین معنی که در شرایطی که اول به مدل پاداش مواد و سپس تنبیه داده شود، مدل رفتار متفاوتی نسبت به حالت عکس (اول تنبیه و سپس مواد) نشان می‌دهد. علت این امر آن است که شرایطی که عملوند دوم رابطه‌ی بیشینه‌گیری، D_{s_t}، بیشتر از عملوند اول آن، $r_{t+1} + \mathrm{V}_{s_{t+1}} - \mathrm{Q}_{s_t}^{a_t} + D_{s_t}$، می‌شود، مقدار ارزش وضعیتی که پس از مصرف مدل به آن می‌رود، $V_{s_{t+1}}$، تاثیری در مقدار سیگنال خطا ندارد. تاثیر

رفتاری این نادیده گرفته شدن را می‌توان در شرایطی که مدل ابتدا مواد دریافت‌می‌کند و سپس تنبیه، مشاهده کرد. این سناریو در شکل ۴-۶ الف نشان داده شده است. با انتخاب عمل فشار دادن اهرم، مدل در ابتدا پاداش مواد را دریافت کرده و سپس به آن شوک داده می‌شود. در صورت انتخاب هر عمل دیگری، مدل یک پاداش نزدیک صفر دریافت کرده، و مواد دریافت نمی‌کند.

فرض کنیم که مقدار تنبیه بسیار بزرگ است و بنابراین ارزش وضعیت V_{S_1} بسیار کم است و لذا داریم $|\mathrm{V}_{S_1}| \ll 0$. بنابراین:

$$r_{t+1} + \mathrm{V}_{s_{t+1}} - \mathrm{Q}_{s_t}^{a_t} + D_{s_t} < D_{s_t} \qquad (۱۳-۴)$$

با توجه به رابطه‌ی بالا مقدار عمل فشار دادن اهرم (PL) همواره با عبارت $D_{s_t} - \rho_t - \kappa_t$ به‌روز خواهد شد. مقدار ارزش هر عمل دیگر (Other) به وسیله‌ی سیگنال خطای عادی به‌روز شده و مقدار آن برابر $R_0 - \rho_t - \kappa_t$ خواهد بود. از آنجایی که مقدار عمل دیگر برابر صفر است، خواهیم داشت $D_{s_t} > E[R_0]$ و بنابراین مقدار ارزش عمل فشار دادن اهرم همواره به وسیله‌ی یک مقدار بزرگتر از هر عمل دیگر به‌روز می‌شود. در نتیجه مدل از ابتدا به طور غیرحساس به مقدارِ تنبیه، عمل فشار دادن اهرم را انتخاب خواهد کرد. لذا پیش‌بینی می‌شود که اگر در ابتدا پاداش مواد داده شود و سپس تنبیه، مدل مستقل از اندازه‌ی تنبیه و طول مدت مصرف، از خود رفتار غیرحساس به تنبیه نشان می‌دهد. این نوع رفتار مدل مربوط به تاخیر در تنبیه و یا مقدار زیاد پاداش مواد نیست و تنها به این خاطر است که در شرایطی که عملوند دوم عمل بیشینه‌گیری زیاد است، ارزش وضعیت‌های پس از مصرف مواد، ارزش وضعیت مصرف مواد را تحت تاثیر قرار نمی‌دهند.



شکل ۴-۶ دو سناریوی مختلف به منظور بررسی تاثیر تقدم و تاخر پاداش و تنبیه. (الف) مدل ابتدا پاداش مواد را دریافت کرده، سپس تنبیه می‌شود. مدل مستقل از تعداد دفعات پیشین مصرف و اندازه‌ی پاداش PL را انتخاب می‌کند. (ب) مدل ابتدا تنبیه دریافت کرده و سپس پاداش مواد دریافت می‌کند. در این حالت مدل وابسته به اندازه‌ی تنبیه و تعداد دفعات مصرف پیشین، ممکن است که عمل R_o و یا PL را انتخاب کند.

وضعیت فوق در شرایطی که مدل در ابتدا تنبیه دریافت کند و سپس پاداش مواد، روی نمی‌دهد (شکل ۴-۶ب). این وضعیت همانند شبیه‌سازی بیان شده در بخش ۴-۴-۳ است و بروز رفتار مصرف مواد، وابسته به اندازه‌ی تنبیه و مدت مصرف پیشین مواد است که پیش‌تر شبیه‌سازی آن ارائه شد.

گرچه رفتار مدل در حالت اول تاکنون به صورت مستقیم آزمایش نشده است، با این وجود این پیش-بینی دور از ذهن به نظر می‌رسد. این احتمال وجود دارد که یادگیری ارزش مواد تنها توسط مسیر دوپامینی که در اینجا مدل شده است صورت نپذیرد. این بدان معنا است ارزش وضعیت مصرف مواد از مسیرهای غیردوپامینی تحت تاثیر تنبیه دریافت‌شده پس از مصرف مواد قرار بگیرد. این فرض با مدل‌های یادگیری که مسیرهای غیر از مسیر دوپامینی را در یادگیری موثر می‌دانند، سازگار است [۷۸].

با تحلیل بیشتر علت وجود عملگر بیشینه‌گیری در مدل ردیش می‌توان به راه‌حل‌های دیگری رسید. به یاد داریم که در مدل ردیش وجود عملگر بیشینه‌گیری بدان علت بود که مقدار سیگنال خطا هرگز

به سمت صفر میل نکند. در مدل بیان شده در این بخش، مقدار سیگنال خطا در برابر فعالیت یکنواخت نورون‌های دوپامینی اندازه‌گیری می‌شود و بالارفتن این فعالیت‌ها، مقدار آن بتدریج به سمت صفر میل می‌کند. با این وصف، وجود عملگر بیشینه‌گیری در مدل پیشنهاد شده از پشتوانه‌ی محکمی برخوردار نیست. به علاوه همانطور که گفته شده، در مدل ردیش عملگر بیشینه‌گیری را نمی‌توان حذف کرد، زیرا در این صورت مقدار ارزش مصرف مواد پس از تعداد محدودی به سمت مقدار نهایی خود می‌رسد و مدل در کوتاه‌مدت از خود رفتار غیرحساس به تنبیه نشان می‌دهد. در مدل پیشنهاد-شده در این بخش حذف عملگر بیشینه‌گیری این مشکل را ایجاد نمی‌کند، زیرا همانطور که در بخش ۴-۴-۳ بیان شد، زیاد شدن مقدار ارزش مواد در درازمدت وابسته به وجود عملگر نبوده و تا موقعی که متوسط پاداش حداقل به D_{s_t} نرسد، ادامه پیدا می‌کند. این بدان معناست که در مدل فعلی می-توان عملگر بیشینه‌گیری را بدون وارد شدن خلل به رفتار مطلوب مدل حذف کرد. در این صورت مقدار سیگنال خطا به صورت زیر خواهد بود:

$$\delta_t^c = r_{t+1} + \mathrm{V}_{s_{t+1}} - \mathrm{Q}_{s_t}^{a_t} + D_{s_t} - (\rho_t + \kappa_t) \qquad (۱۴-۴)$$

با اعمال رابطه‌ی فوق، چه پاداش مواد قبل از تنبیه، چه بعد از آن داده شود،رفتار مدل یکسان خواهد بود. بررسی دیگر رفتارهای مدل بیان‌شده در (۱۴-۴) به پژوهش‌های بعدی موکول می‌شود.

۴-۶ بحث

با در نظر گرفتن سه فرض عصب‌شناسی:

۱. فعالیت فازیک نورون‌های دوپامینی، کدکننده‌ی خطای پیش‌بینی هستند،
۲. مواد اعتیادآور باعث تولید مصنوعی دوپامین می‌شوند،
۳. مصرف مزمن مواد اعتیادآور باعث بالارفتن غیرعادی آستانه‌ی پاداش می‌شود،

مدلی برای اعتیاد به کوکائین پیشنهاد شد. مدل پیشنهاد شده علاوه بر انطباق با مصرف غیرحساس به هزینه در اعتیاد مزمن، کاهش ارزش پاداش‌های طبیعی و افزایش رفتار تکانشگرایانه را نیز توضیح می‌دهد. همچنین مدل پیش‌بینی می‌کند که پدیده‌ی بلوکه کردن در مورد پاداش مواد روی دهد. مدل پیشنهادشده بر خلاف مدل ردیش، و نیز گوتکین و همکاران، اختلال در پردازش پاداش‌های

طبیعی را ایجاب می‌کند. همچنین مقدار ارزش وضعیت مصرف مواد به طور نامحدود زیاد نمی‌شود، و همچون مدل گوتکین پس از مصرف دراز مدت، یادگیری به طور کلی از بین نمی‌رود.

مدل‌های حیوانی مصرف مواد نشان می‌دهند که تنها درصد کمی از موش‌ها پس از مصرف مزمن مواد دچار رفتار اجباری جستجو و مصرف مواد می‌شوند [۶۷,۷۹]. این نوع تفاوت فردی و استعداد زیستی در معتاد شدن در مدل فعلی بازنمایی نشده است. با این وجود مدل دارای پارامترهای آزادی است که مقدار آنها تا حدودی به خصلت‌های فردی بستگی دارد. مثلا همانطور که نشان داده شد، بروز رفتار غیرحساس به تنبیه وابسته به مقدار D_{st} است و به ازای مقادیر کم این متغیر، رفتار جستجوی اجباری بروز نمی‌کند. برخی از تفاوت‌های فردی را می‌توان توسط مقدار این متغیرها توضیح داد. با این وجود تفاوت‌های فردی دیگری نیز که پایه ارثی دارند در معتادشدن دخیل هستند، که پرداختن به آنها به کارهای آینده واگذار می‌شود.

علاوه بر بخش‌هایی از نظام عصبی که مدل‌سازی شد، شواهد نشان می‌دهد بخش‌های دیگری از مغز نیز مانند آمیگدال[۱]، غشایی پیش‌جلویی[۲] و پالیدوم[۳] مورد تاثیر مواد قرار گرفته، در بروز رفتار معتادانه دخیل هستند. به عنوان مثال شواهد نشان می‌دهد که پس از مصرف مزمن مواد، کنترل رفتار به نحو غیرعادی از بخش‌های غشایی که مسئول تصمیم‌گیری رفتار هدف‌مدار[۴] هستند، به مغز میانی که مسئول رفتارهای عادتی[۵] است انتقال پیدا می‌کند. با کمی مسامحه می‌توان رفتار هدف‌مدار را با یک مدل یادگیری تقویتیِ مبتنی بر مدل، و رفتار عادتی را توسط یک یادگیری تقویتی مستقل از مدل الگوسازی کرد. مدل بیان شده در این فصل را می‌توان به عنوان مدلی از رفتار عادتی در نظر گرفت، که انتقال کنترل رفتار به طور غیرعادی از بخش‌های هدف مدار به آن منتقل شده است. همانند قبل، به منظور مدل‌سازی این پدیده (انتقال غیرعادی کنترل به بخش‌های عادتی)، در ابتدا باید الگویی برای انتقال کنترل بین بخش‌های هدف‌مدار و عادتی در افراد سالم ارائه شود و سپس تاثیر مواد بر این انتقال مورد مطالعه قرار گیرد. قدم اول را می‌توان بر اساس مدل پیشنهادشده در [۸۰] برداشت که در

[1] Amygdale
[2] Prefrontal cortex
[3] Pallidum
[4] Goal-based
[5] Habitual

آن الگویی برای انتقال کنترلِ رفتار بین بخش‌های هدف‌مدار و عادتی ارائه شده است. اساس مدل به این صورت است که هر دو سیستم (هدف‌مدار و عادتی) با استفاده از الگوی یادگیری تقویتی بیزی، تخمینی از ارزش پاداش عایدی از انجام یک عمل ارائه می‌کنند. سیستم هدف‌مدار از یادگیری تقویتی بیزی مبتنی بر مدل (که بر اساس روش تکرار ارزش کار می‌کند) پیش‌بینی انجام داده و سیستم عادتی با استفاده از یک مدل یادگیری تقویتی بیزی مستقل از مدل عمل تخمین را انجام می‌دهد. در هر لحظه از زمان، کنترل رفتار را سیستمی که تخمین دقیق‌تری از نتیجه‌ی عمل داشته باشد به عهده می‌گیرد. با این تلقی از یک مدل سالم، در مورد اعتیاد باید تخمین ارزش‌ها توسط بخش هدف مدار به طور غیرعادی نادقیق شده و این موجب شود که رفتار به طور غیرعادی به بخش‌های عادتی منتقل شود؛ یا بر عکس تخمین‌ها در نظام عادتی باید بیش از حد قطعی باشد. در حال حاضر شاهدی بر هیچ یک از این ادعاها در دسترس نیست، از طرفی به لحاظ شهودی نیز منطقی به نظر نمی‌رسند. بررسی بیشتر این امر به کارهای آینده واگذار می‌شود.

منحرف شدن سیگنال خطا از مقدار اصلی خود، باعث می‌شود که مقادیر تخمین زده‌شده برای ارزش پاداش‌های طبیعی و مواد در درازمدت افت کنند. به لحاظ محاسباتی، این پدیده همگرا شدن مقادیر را با مشکل روبرو نمی‌کند. زیرا مقادیر ارزش وضعیت‌ها در مدل یادگیری تفاضل زمانی متوسط پاداش نسبت به یک وضعیت دلخواه که ارزش آن صفر فرض می‌شود اندازه‌گیری می‌شوند [۱۹]. بنابراین کم شدن ارزش وضعیت‌ها منجر به واگرایی مقادیر آنها نمی‌شود.

۴-۷ خلاصه

در این فصل به معرفی یک مدل عصبی-محاسباتی برای اعتیاد به کوکائین پرداختیم. برای این کار، در ابتدا ساختار مدل را بر اساس شواهد عصبی طراحی کرده، سپس در خلال شبیه‌سازی در مورد رفتار آن در شرایط مختلف بحث کردیم. پس از آن به ارائه‌ی پیش‌بینی پرداخته و شرایط درست بودن و اشتباه بودن آن را مورد بررسی قرار دادیم. در انتها نیز به طور اجمالی در مورد مدل بحث کردیم. در جدول ۴-۲ مقایسه بین مدل‌های محاسباتی مختلف اعتیاد ارائه شده است.

جدول ۴-۲ مدل‌های دارویی و عصبی محاسباتی-اعتیاد و مقاسیه آن‌ها. √ به معنای توانایی توضیح. – به معنای عدم توانایی توضیح و ؟ به معنای نامشخص بودن وضعیت است.

	مدل پیشنهاد شده	مدل ردیش	مدل گوتکین	مدل آهمد و کوب	مدل نرمان و تسیبولسکی
رفتار غیرحساس به تنبیه (Compulsive Drug Seeking)	√	√	√	؟	؟
کاهش انگیزه برای پاداش‌های طبیعی	√	–	–	–	–
بازگشت (Reinstatement)	–	–	–	–	–
تکانشگری	√	–	–	–	–
نرخ پاسخ در شرایط همراه با سیری (Response rate under satiety effect)	–	–	–	√	√
نرخ پاسخ در شرایط بدون سیری (Reponse rate without satiety effect)	–	–	–	–	–
شروع اعتیاد (Aquisition)	؟	؟	؟	–	√
ادامه‌ی اعتیاد (Maintenance)	؟	؟	؟	√	–
کاهش توانایی کوکائین در زیاد کردن میزان دوپامین پس از مصرف دراز مدت	√	–	–	–	–
بالارفتن سطح پاداش پس از مصرف دراز مدت	√	–	–	√	–
انتقال غیرعادی کنترل به تصمیم‌گیری عادتی	–	–	–	–	–
توضیح پدیده‌ی بلوک کردن	√	–	؟	؟	؟
قابلیت توسعه به شرایط غیر آزمایشگاهی	√	√	√	–	–
ترکِ رفتارِ پاسخ برای دریافت مواد	√	√	–	؟	؟

بخش دوم:

مدل‌سازیِ شناختیِ اعتیاد

فصل ۵

مدل‌سازی شناختی آزمونِ قمار آیوا

۵-۱ مقدمه

همان‌طور که در فصل اول بیان شد، اعتیاد را می‌توان به عنوان یک رفتار اجباری به سمت مصرف مواد، درحالیکه فرد معتاد از آثار زیان‌بار آن در آینده آگاه است، معرفی نمود. اینگونه اختلال‌های تصمیم‌گیریِ ناشی از اثرهای درازمدت مداخله‌های مواد بر نظام تصمیم‌گیری در مغز است؛ و اغلب تاثیر آنها به رفتار فرد معتاد در برابر ماده‌ی اعتیادآور محدود نمانده و به تصمیم‌گیری‌های روزمره وی تعمیم می‌یابد. بر این اساس، آزمون‌های ارزیابی شناختی[۱] سعی می‌کنند با شبیه‌سازی وجوهی از تصمیم‌گیری در زندگی روزمره، چگونگی تصمیم‌گیریِ افراد را مورد ارزیابی و اندازه‌گیری قرار دهند. هدف از ارزیابی شناختی در درجه‌ی اول سنجش توانایی شناختی افراد و در درجه‌ی دوم کشف علت یک نارسایی شناختی در نمونه‌ی مورد مطالعه است.

آزمون قمار آیوا[۲] از متداول‌ترینِ آزمون‌ها برای سنجش تصمیم‌گیری افراد تحت شرایط مخاطره‌آمیز و غیرقطعی است [۸۱]. آزمون قمار یک محیط تصمیم‌گیری واقعی را برای آزمایش دهنده شبیه‌سازی

---

[۱] Cognitive Assessment Tasks
[۲] Iowa Gambling Task

می‌کند. مطالعه‌های پیشین حاکی از آن است که این آزمون قادر است گروه افراد معتاد به سوء مصرف مزمن مواد اعتیادآور را از افراد سالم بر حسب کارایی تصمیم‌گیری تفکیک کند (اختیاری و همکاران، ۱۳۸۳). نشان داده شده که به طور کلی افراد معتاد در آزمون کارایی کمتری نسبت به گروه کنترل دارند.

به طور سنتی در پژوهش‌های بالینی و ارزیابی شناختی، فرایند ارزیابی شناختی با در نظر گرفتن یک آزمون به هدف سنجش یک توانایی شناختیِ خاص شروع می‌گردد. در مرحله‌ی بعد سعی می‌شود که شاخص‌های مناسبی از عملکرد فرد در آزمون استخراج شود. شاخص مطلوب شاخصی است که بر اساس تحلیل‌های آماری بین مقدار آن و عضویت فرد در گروه شاهد[1] یا بیمار همبستگی کافی وجود داشته باشد. در صورت کشف چنین شاخصی می‌توان از آزمون و شاخص معرفی شده به عنوان یک آزمون ارزیابی کننده‌ی تواناییِ تصمیم‌گیری افراد بهره جست.

گرچه فرایند ذکرشده در بالا اطلاعات ارزشمندی در مورد سازوکار تصمیم‌گیری در گروه بیمار فراهم می‌کند، لیکن قادر به ارائه‌ی جوابی در خصوص چرایی رفتار مشاهده‌شده در یک آزمون نیست. به عنوان مثال این پرسش که چرا گروه معتاد نسبت به گروه شاهد امتیاز کمتری در آزمون قمار کسب کرده‌اند از طریق دسته‌بندی آنها در دو گروه به وسیله‌ی شاخص عملکرد آنها، قابل پاسخ نیست. یک روش برای پاسخ دادن به این گونه پرسش‌ها بهره‌گیری از روش مدل‌سازی شناختی[2] است.

مدل‌سازی شناختی [۸۲] رویکردی محاسباتی برای توصیف فرایندهای شناختیِ درگیر در یک عمل شناختی است. مدل‌سازی شناختی خود گونه‌ای از مدل‌سازیِ محاسباتی است که بر اساس این فرض استوار است که سازوکارهای عصبی دخیل در یک عمل شناختی هر یک محاسبه‌ای را بر روی داده-های ورودی از حسگرها یا خروجی‌های بخش‌های مختلف مغز پیاده‌سازی می‌کنند. مثلا در مورد تصمیم‌گیری، این مجموعه از محاسبه‌ها هدایت‌کننده‌ی انتخاب‌های یک تصمیم‌گیرنده است. مدل-سازی محاسباتی این مجموعه از فعالیت‌ها را توسط روابط ریاضی توصیف می‌کند. نقطه‌ی تمایز اصلی این روش با روش‌های تحلیل آماری در آن است که در مدل‌سازی شناختی، یک عمل شناختی بر

[1] Control group
[2] Cognitive modeling

اساس فرایندهای شناخته شده‌ی مغزی توضیح داده می‌شود. مثلا در توضیح چگونگی نحوه‌ی تصمیم‌گیری از عناصری مانند حافظه، ارزیابی پاداش، توجه و غیره استفاده می‌شود. این ویژگی این امکان را فراهم می‌کند که بتوان بین یک رفتار بیمارگونه‌ی مشاهده‌شده و زیرساخت‌های روانی (و در صورت امکان عصبیِ) آن ارتباط برقرار کرد. تفاوت این روش با الگوی پیشین تحلیل رفتار در یک آزمون شناختی در آن است که این روش تنها به استخراج شاخصی از نحوه‌ی تصمیم‌گیری فرد بسنده نکرده و کل فرایند تصمیم‌گیری را مدل‌سازی می‌کند. تفاوت اصلی مدل‌سازی شناختی با روش‌های مدل‌سازی مفهومی در این است که در مدل‌سازی شناختی از زبان ریاضیات به منظور توصیف فرایند استفاده می‌شود. استفاده از زبان ریاضیات امکان ارزیابی کمّی انطباق یک مدل با واقعیت را فراهم می‌کند.

در پژوهش ارائه شده در این فصل، در قدم اول عملکرد تصمیم‌گیری دو گروه معتاد و شاهد به شرحی که در بخش بعد آمده است توسط آزمون قمار ایوا مورد سنجش قرار گرفت. این بخش از کار توسط مرکز ملی مطالعات اعتیاد ایران انجام شد. در قدم بعد منشاء رفتار هر دو گروه شاهد و معتاد توسط مدل‌سازی شناختی رفتار آنها مورد بررسی قرار گرفت. مدل‌های استفاده شده بر اساس تئوری یادگیری تقویتی طراحی شده‌اند که منطق تصمیم‌گیری در آنها در بخش بعد بیان شده است. در انتها نیز نتایج حاصل از این مطالعه ارائه شده است.

۵-۲ روش‌ها و ابزارها

۵-۲-۱ آزمون قمار آیوا: ابزاری برای ارزیابی شناختی تصمیم‌گیری

آزمون قمار آیوا به عنوان چارچوبی برای ارزیابی بسیاری از اختلال‌های تصمیم‌گیریِ انسان به کار گرفته شده‌است. در نسخهٔ $ABCD$ی این آزمون که در این پژوهش مورد استفاده قرار گرفته است، چهار دسته کارت ۶۰تایی پیش‌روی آزمون‌دهنده قرار می‌گیرد. آزمون‌دهنده هر بار یک کارت را از یکی از چهار دسته انتخاب می‌کند. پس از هر انتخاب، میزان برد یا باخت به او اطلاع داده می‌شود. آزمون‌دهنده در کل ۱۰۰ انتخاب دارد و باید سعی کند که در طی انتخاب‌ها به بیشترین میزان سود خالص دسترسی پیدا کند. کارت‌ها به دو دسته‌ی سودآور (C,D) و ضرر‌ده (A,B) تقسیم‌بندی می-

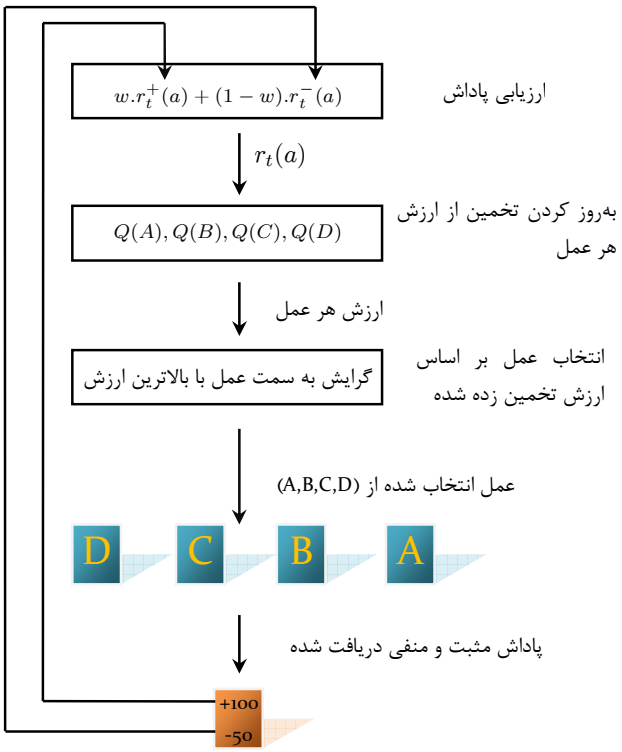
جدول ۵-۱خصوصیت جمعیتی آزمودنی‌ها

	گروه‌ها	گروه شاهد	گروه معتاد
شاخص‌های دموگرافیک	تعداد	۱۳۰ نفر	۲۱۷ نفر
	سن (سال)	۳۰/۲۵ ± ۸/۷۷	۲۹/۸۷ ± ۷/۶۰
	تحصیلات (سال)	۱۲/۰۳ ± ۳/۷۷	۱۱/۴۹ ± ۳/۰۰

شوند. کارت‌های سودآور اگر چه میزان سودی که نصیب فرد می‌کنند کم است، ولی با توجه به کمتربودن باخت نسبت به سود به طور مجموع سودآور هستند. کارت‌های ضرر‌ده نسبت به کارت‌های سوده، سود بیشتری نصیب عامل می‌کنند، ولی در کل با توجه به بیشتربودن میزان ضرر آنها نسبت به سود، ضرر‌ده هستند. میزان سود (در دسته کارت‌های سوده) و ضرر (در دسته کارتهای ضرر‌ده) به مرور زمان افزایش می‌یابد. گرچه کارت‌های A و B بطور میانگین دارای سود برابر هستند، لیکن تعداد کارت‌هایی که در دسته‌ی A باخت به همراه دارند بیشتر، ولی میزان باخت هر کدام کمتر است. در مقابل، در کارت‌های دسته‌ی B، تعداد کارت‌های دارای باخت کمتر است، ولی میزان باخت هر کارت بیشتر است. همین قانون در مورد کارت‌های دو دسته‌ی C و D نیز صادق است. در انتهای آزمون، امتیاز خالص آزمون‌دهنده به صورت $(C+D)-(A+B)$ محاسبه می‌شود. در این پژوهش از نسخه فارسی این آزمون استفاده شد (اختیاری و همکاران، ۱۳۸۳).

۵-۲-۲ آزمودنی‌ها

گروه معتادِ مورد آزمون، شامل ۲۱۷ نفر از افراد معتاد به مواد افیونی (بر اساس معیارهای *DSM-IV*) که به منظور شرکت در پروتکل‌های درمانی به مرکز ملی مطالعات اعتیاد ایران مراجعه کرده‌اند، می‌باشد. گروه شاهد شامل ۱۳۰ نفر بدون سابقه‌ی مصرف مواد (به‌جز سیگار) بوده است. گروه شاهد از بستگان افراد گروه معتاد انتخاب شده‌اند که هماهنگی قابل‌قبولی میان خصوصیات جنسیتی، سن و سطح تحصیلات در این گروه با گروه معتادان وجود دارد. خصوصیات جمعیتی این دو گروه درجدول ۵-۱ ارائه شده است.



شکل ۵-۱ معماری کلی مدل‌های استفاده‌شده به منظور ارزیابی شناختی. مدل‌ها از سه بخش تشکیل شده‌اند. در بخش اول پاداش و ضرر دریافت‌شده مورد ارزیابی قرار می‌گیرند و وزن‌دهی می‌شوند (مشترک بین همه مدل‌ها). در بخش دوم از حاصل مرحله‌ی قبل به منظور یادگیری ارزش هر دسته کارت استفاده می‌شود (متفاوت بین مدل‌های مختلف). در مرحله‌ی آخر بر اساس ارزش تخمین زده شده از هر دسته کارت، از یکی از دسته‌ها کارت انتخاب می‌شود (متفاوت بین مدل‌ها).

۵-۲-۳ مدل‌های تصمیم‌گیری

در این بخش از مقاله به توضیح مدل‌های مختلف تصمیم‌گیری که در این پژوهش مورد استفاده قرار گرفته‌اند، می‌پردازیم. در کل دوازده خانواده‌ی مختلف از مدل‌ها مورد استفاده قرار گرفت. تمامی مدل‌ها از سه بخش اصلی تشکیل شده‌اند (شکل ۵-۱).

بخش اول در همه‌ی مدل‌ها مشترک است و مسئول ارزیابی پاداش و ضرر دریافتی پس از انتخاب یک کارت است. پس از انتخاب یک کارت مانند کارت a (a می‌تواند هریک از کارت‌های A, B, C یا D باشد) عامل پاداش مثبت $r_t^+(a)$ و یا پاداش منفی $r_t^-(a)$ یا هر دو را دریافت می‌کند. مثلا اگر پس از برداشتن یک کارت از دسته کارت A، آزمون‌دهنده عددهای 100+ (میزان برد) و 200- (میزان

باخت) را در پشت کارت مشاهده کند، 100+ همان $r_t^+(a)$ بوده و $r_t^-(a)$ برابر 200- می‌باشد. در بخش ارزیابی پاداش، پاداش‌های مثبت و منفی دریافت شده به صورت خطی توسط یک پارامتر به نام پارامتر ارزش‌گذاری[1]، که آن را با w نشان می‌دهیم با یکدیگر ترکیب شده و از حاصل آن در مرحله‌ی بعد استفاده می‌شود. پارامتر w چگونگی این ترکیب را تعیین می‌کند و مقدار آن در بازه‌ی صفر و یک قرار دارد. مقادیر w نزدیک به یک بیانگر این است که مدل در یادگیری تنها مقادیر پاداش‌های مثبت دریافت شده ($r_t^+(a)$) را لحاظ می‌کند و نسبت به مقادیر ضررها ($r_t^-(a)$) بی‌توجه است. از طرف دیگر مقادیر نزدیک به صفر بدان معنا است که عامل تنها ضررهای دریافتی را لحاظ می‌کند و نسبت به پاداش‌های مثبتِ دریافتی بی‌توجه است. در اینجا ما حاصل ترکیب پاداش‌های مثبت و منفی را پاداش دریافت‌شده می‌نامیم.

بخش دوم بخش یادگیرنده است که از پاداش‌های دریافت شده به منظور یادگیری استفاده می‌کند. خروجی مرحله‌ی یادگیری تخمین ارزش هر دسته از کارت‌ها است که این تخمین‌ها درشکل ۵-۱ به وسیله $Q(C), Q(B), Q(A)$ و $Q(D)$ نشان داده شده‌اند. مثلا مقدار $Q(D)$ در سی‌امین انتخاب مدل، نشان‌دهنده‌ی انتظار مدل از آنچه در صورتِ انتخابِ کارت از دسته D عایدش می‌شود، است. بر خلاف بخش پیشین که در همه‌ی مدل‌های استفاده شده مانند هم است، این بخش در مدل‌های مختلف متفاوت است و مدل‌ها از روش‌های مختلفی به منظور یادگیری ارزش هر دسته‌کارت استفاده می‌کنند. الگوهای ارائه شده برگرفته‌شده از مطالعه قبلی انجام شده بر روی داده‌های آزمون قمار در بیماران با ضایعات مغزی در ناحی میانی تحتانی قشر پرده فرونتال[2] است [۸۳]. در کل شش الگوی مخلتف یادگیری استفاده شده است که می‌توان آنها را به دو دسته تقسیم کرد.

دسته‌ی اول مدل‌ها برای یادگیری ارزش هر دسته از کارت‌ها، از مقدار پاداش دریافت‌شده استفاده می‌کنند. مثلا در یکی از مدل‌های این دسته برای پیش‌بینی ارزش هر دسته از پاداش‌های دریافت‌شده قبلی، میانگین‌گیری می‌شود. در این صورت مقدار ارزش دسته‌ی کارت D در انتخاب مثلا سی‌ام برابر

[1] Valence parameter
[2] Ventro medial prefrontal cortex

است با میانگین پاداش‌های کسب شده در انتخاب‌های قبلی از این دسته کارت. این دسته از مدل‌ها را "یادگیری مبتنی بر مقدار" می‌نامیم.

تفاوت مدل‌های دسته‌ی دوم با مدل‌های دسته‌ی اول در این است که این مدل‌ها به مقدار پاداش دریافت شده بی‌توجه هستند و تنها به این که حاصل پاداش دریافت شده مثبت یا منفی است، توجه دارند. مثلا پس از انتخاب یک کارت، پاداش دریافت شده برابر 100+ یا 500+ باشد، تاثیر یکسانی در پیش‌بینی‌های بعدی در مورد ارزش این دسته‌کارت دارند. ولی اگر مثلا مقدار پاداش دریافت‌شده برابر 50- باشد (باخت یا تنبیه) در این صورت تاثیر متفاوتی در پیش‌بینی‌های بعدی گذاشته و مقدار ارزش دسته کارت را کم می‌کند. ولی مقدار کم شدن این ارزش در صورتی که مدل پاداش 600- یا 50- را دریافت کند، تفاوتی نخواهد کرد. ما الگوهای این دسته را "یادگیری مبتنی بر تعداد" می‌نامیم.

واضح است که نادیده گرفتن مقدار پاداش، و یادگیری تنها بر اساس مثبت یا منفی بودن آن، عامل را به تخمین اشتباه از ارزش کارت‌ها می‌رساند. لیکن هدف از در نظر گرفتن مدل‌های مختلف این است که بررسی شود کدامیک با رفتار مشاهده‌شده در آزمون بیشتر منطبق است. مثلا اگر دسته‌ی دوم مدل‌ها بهتر از دسته‌ی اول بر رفتار یک فرد در آزمون قمار برازش پیدا کرد، می‌توان نتیجه گرفت که فرد بیشتر از اینکه تحت تاثیر مقدار پاداش و ضرر قرار داشته باشد، تحت تاثیر دفعات قرار دارد. هر یک از دو دسته مدل‌های یادگیری در درون خود دارای چند نوع مختلف هستند که از شرح آنها و روابط ریاضی هر یک در اینجا صرف‌نظر شده و خواننده محترم می‌تواند به پیوست الف مراجعه کند.

چگونگی یادگیری در الگوهای یادگیری به وسیله‌ی پارامتر نرخ یادگیری که آن را با γ نشان می‌دهیم کنترل می‌شود. این پارامتر در بازه‌ی $[0, 1]$ قرار دارد و اگر مقدار آن در یک مدل نزدیک به یک باشد بیانگر آن است که مدل تنها به تجربه‌های جدید خود برای تخمین ارزش هر دسته کارت توجه می‌کند و به تجربه‌های قبلی خود اهمیتی نمی‌دهد. از طرف دیگر مقادیر نزدیک به صفر آن نشان می‌دهد که اثر تجربیات جدید در یادگیری ارزش هر دسته از کارت‌ها کم است و ارزش‌ها به کندی از تجربیات جدید تاثیر می‌گیرند.

جدول ۵-۲ نتایج آزمون قمار بر اساس جمع و تفریق دفعات انتخاب از کارت‌های A, B, C یا D در طی آزمون (مجموعه‌ی یکصد انتخاب)

	گروه‌ها	گروه شاهد	گروه معتاد
نمرات آزمون قمار	(C+D) – (A+B)	۶/۸۸ ± ۲۸/۴۷	-۲/۳۹ ± ۲۶/۱۰
	(B+D) –(A+C)	۱۹/۶۵± ۲۱/۵۵	۱۷/۵۲ ± ۲۱/۸۵

مرحله‌ی سوم، انتخاب یک کارت بر اساس تخمین‌های محاسبه شده از ارزش هر دسته‌کارت است. دو الگوی مختلف انتخاب عمل روش ϵ-حریصانه و بیشینه نرم در این پژوهش مورد استفاده قرار گرفته است. از ترکیب شش الگوی یادگیری با دو الگوی انتخاب عمل دوازده مدل حاصل می‌شود که در این مطالعه هرکدام جداگانه بر رفتار مشاهده شده از هر گروه، برازش داده و کیفیت برازش آنها با یکدیگر مقایسه شده است. به منظور کشف اینکه کدام مدل بهتر با رفتار یک گروه منطبق بوده است از معیار اطلاعات بیزی[1] استفاده شده است، که شرح آن در بخش ضمیمه آورده شده است. برای هر گروه به طور جداگانه مقادیر پارامترهای آزاد مدل، مانند w، γ، تخمین زده شد. منظور از تخمین پارامتر شناسایی مقداری از پارامتر است که وقتی مدل با آن مقدار شبیه‌سازی می‌شود، رفتاری نزدیک‌تر به رفتار مشاهده شده از گروه مورد نظر بروز پیدا می‌کند.

در بخش بعدی نتایج حاصل از آزمون و مدل‌سازی آنها شرح داده شده است.

۵-۳ نتایج

نتایج حاصل از آزمون قمار در دو گروه شاهد و معتاد در جدول ۵-۲ آمده است. گرچه نتایج گروه شاهد به صورت معناداری ($p < 0.05$) بهتر از گروه افراد معتاد است، اما در مقایسه با نتایج مطالعات غربی انجام‌شده که حداقل نمرهٔ ۱۰ را معرف عملکرد افراد سالم می‌دانند [۸۴] این نتایج نشانگر ضعف گروه شاهد در تصمیم‌گیری است.

از میان انواع روش‌های یادگیری تقویتی، مدل بهینه برای هر دو گروه، مدل یادگیری مبتنی بر تعداد به همراه روش انتخاب عمل بیشینه نرم می‌باشد. همان‌طور که در بخش قبل گفته شد در مدل

[1] Bayesian information criterion

یادگیری مبتنی بر تعداد، مقادیر برد و باخت در یادگیری نقشی ندارند؛ و به جای آنها از تعداد برد و باخت به منظور یادگیری استفاده می‌شود. بهتر بودن توانایی این مدل در توصیف رفتار هر دو گروه بیان کننده‌ی آن است که کم بودن امتیاز خالص هر دو گروه به علت نادیده گرفتن مقادیر برد و باخت و توجه به تعداد برد و باخت و در نتیجه انتخاب از کارت‌های ضرر‌ده بوده است.

مقادیر بهینه‌ی پارامترها به ازای این مدل بصورت زیر می‌باشند:

$$P^*_{control} = (\hat{\beta} = 0.70, \hat{w} = 0.35, \hat{\gamma} = 0.20) \quad (۱-۵)$$

$$P^*_{addict} = (\hat{\beta} = 0.50, \hat{w} = 0.15, \hat{\gamma} = 0.35) \quad (۲-۵)$$

که در آن، $\hat{\beta}$ میزان بهره‌برداری، $\hat{w}$ پارامتر ارزش‌گذاری و $\hat{\gamma}$ نرخ یادگیری است. بر اساس مدل بهینه برای هر گروه می‌توان گفت:

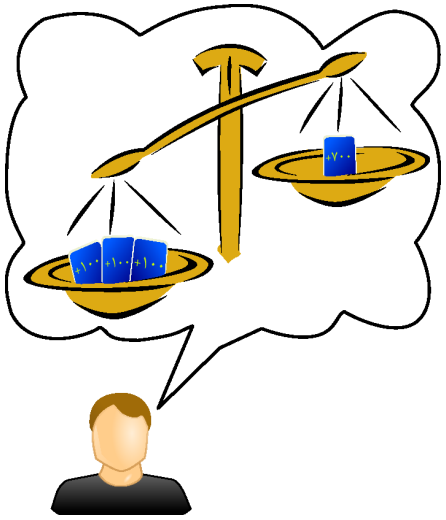
۱- با توجه به اینکه مدل یادگیریِ بر اساس تعداد دفعات، توصیف‌کننده‌ی بهتری برای رفتار هر دو گروه نسبت به مدل یادگیری بر اساس مقدار است، می‌توان نتیجه گرفت که هر دو گروه بیش از آنکه به مقادیر پاداش و تنبیه (برد و باخت) حساس باشند، به تعداد دفعات هر یک حساسیت دارند (شکل ۵-۲).

۲- با توجه زیادتر بودن مقدار پارامتر $\hat{\beta}$ در گروه شاهد، می‌توان گفت که گروه شاهد نسبت به گروه معتاد، تمایل بیشتری به انتخاب اعمال با ارزش انتظاری بیشتر دارد.

۳- با توجه به بزرگ‌تر بودن مقدار پارامتر $\hat{w}$ در گروه معتاد، می‌توان نتیجه گرفت که رفتار آسیب‌گریزی در گروه معتاد، بیش از گروه شاهد می‌باشد.

۴- با توجه به بزرگ‌تر بودن نرخ یادگیری در گروه شاهد می‌توان گفت که گروه شاهد، اهمیت بیشتری برای تجربه‌های جدید کسب شده قائل است.

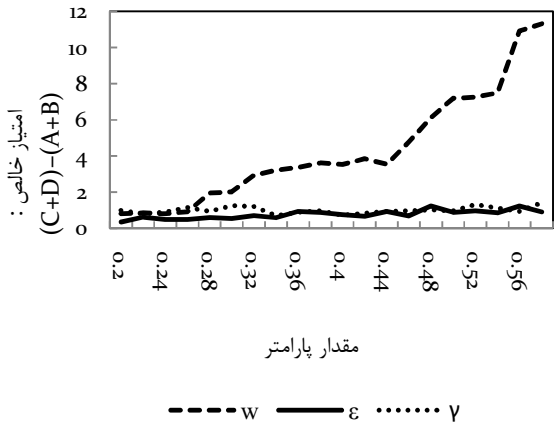
از آنجایی که هر دو گروه شاهد و معتاد، تعداد پاداش و تنبیه را بیش از مقادیر آنها در تصمیم‌گیری لحاظ می‌کنند، باید به دنبال عامل متمایزکنندهٔ دیگری گشت که اختلاف در سطح کارایی دو گروه را



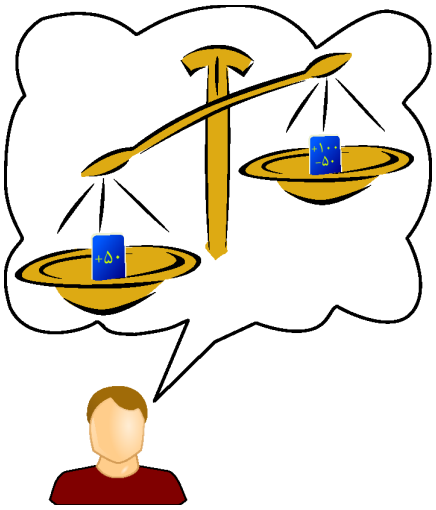
شکل ۵-۲ عامل تحت تاثیر دفعات ضرر و پاداش. برای عامل دریافت سه بار پاداش ۱۰۰ واحدی با ارزش‌تر از دریافت یک پاداش ۷۰۰ واحدی است.

توضیح دهد. شکل ۵-۲ کاراییِ مدل یادگیری مبتنی بر تعداد و روش انتخاب عمل بیشینه نرم را به ازای انحراف هر یک از پارامترها از نقطه‌ی بهینه‌ی P^*_{addict} نشان می‌دهد.

همانطور که در شکل ۵-۳ مشاهده می‌شود، کارایی مدل حساسیت زیادی نسبت به تغییرات پارامتر ارزش‌گذاری (w) از خود نشان می‌دهد و هنگامی که مقدار این پارامتر به مقدار متناظر خود در مدل بهینه گروه شاهد می‌رسد، کارایی مدل بهینه گروه معتاد به کارایی مدل بهینه گروه شاهد می‌رسد. بنابراین می‌توان نتیجه گرفت که انحراف افراد معتاد به سمت رفتار آسیب‌گریزی در کنار تاثیرپذیری زیاد از تعداد دفعات پاداش –به‌جای مقادیر آن- عوامل توضیح دهنده‌ی کارایی پایین گروه معتاد هستند (شکل ۵-۴).



شکل ۵-۳ کارایی مدل بهینه برای افراد معتاد به ازای مقادیر مختلف پارامترها در اطراف بردار P^*_{addict}



شکل ۵-۴ عامل با گرایش به آسیب گریزی. برای یک عامل دریافت یک پاداش ۵۰ واحدی ارزشمندتر از یک پاداش ۱۰۰ واحدی و یک ضرر ۵۰ واحدی است.

۵-۴ بحث

در این پژوهش در قدم اول دو گروه معتاد و شاهد با استفاده از آزمون قمار آیوا مورد ارزیابی شناختی قرار گرفتند (توسط مرکز ملی مطالعات اعتیاد ایران). نتایج حاصل حاکی از ضعف تصمیم‌گیری هر دو گروه در آزمون قمار و ضعیفتر بودن گروه معتاد نسبت به گروه شاهد در این آزمون بود. همانطور که

در بخش مقدمه ذکر شد عدم توانایی افراد وابسته به مواد اعتیادآور در آزمون قمار در مطالعات پیشین نیز مشاهده شده است.

به منظور بررسی علت این مشاهده‌ها از روش مدل‌سازی شناختی استفاده شد. بدین منظور انواع مختلفی از یادگیری تقویتی بر رفتار هر دو گروه برازش داده شد و توانایی توصیف هر الگو مورد بررسی قرار گرفت. نتایج بدست آمده بیان کننده‌ی این است که هر دو گروه تحت تاثیر تعداد دفعات برد و باخت بوده و نسبت به مقدار آنها بی‌توجه هستند. این عامل می‌تواند توضیح‌دهنده‌ی ضعف هر دو گروه در آزمون باشد. از طرف دیگر تحلیل حساسیت مدل بهینه برای گروه معتاد این واقعیت را آشکار کرد که افراد معتاد نسبت به افراد سالم گرایش به سمت آسیب‌گریزی دارند و این عامل توضیح دهنده‌ی عملکرد ضعیف‌تر افراد معتاد نسبت به افراد سالم در آزمون است.

در مطالعات پیشین از روش مدل‌سازی شناختی به منظور تحلیل علت عدم کارایی معتادان در آزمون قمار استفاده شده است [۸۵,۸۶]. در این مطالعه‌ها تنها مدل مبتنی بر مقدار پاداش بر رفتار معتادان برازش داده شده است (مدل مبتنی بر دفعات بر رفتار گروه‌ها برازش داده نشده است). نتایج به دست آمده در این مطالعات بیانگر این است که گرایش معتادان به پاداش جویی علت اصلی تفاوت رفتار آنها با گروه شاهد بوده است. در واقع این مطالعات علت عملکرد ضعیف معتادان را نه در کاستی حافظه یا یادگیری، بلکه در عدم تعادل در توجه به پاداش و ضرر یافته‌اند. از این منظر یافته‌ی این مطالعه‌ها با یافته‌های این پژوهش منطبق است: در پژوهش حاضر نیز علت اصلی تفاوت عملکرد گروه معتاد با گروه شاهد، تفاوت در میزان توجه به پاداش و ضرر بوده است. از منظر انتخاب عمل، در مطالعه‌های پیشین نیز همانند این پژوهش انتخاب عملِ معتادان نسبت به گروه شاهد کمتر تحت تاثیر ارزش پیش‌بینی شده برای هر عمل قرار دارد [۸۶]. نقطه‌ی تفاوت یافته‌های این مطالعه با مطالعه‌های پیشین در این است که در این مطالعه علت کارایی پایین معتادان را گرایش به آسیب گریزی دانسته است. این در حالی است که در مطالعه‌های قبلی علت اصلی عدم آسیب‌گریزی و گرایش به پاداش جویی تشخیص داده شده است.

علت تفاوت یافته‌های این پژوهش با مطالعات پیشین را می‌توان در موارد زیر خلاصه کرد. نکته‌ی اول آنکه در مطالعات پیشین از خانواده‌ی محدودتری (نسبت به مطالعه‌ی فعلی) از مدل‌ها برای مدل‌سازی رفتاری استفاده شده است. مدل‌های استفاده شده در پژوهش‌های پیشین شامل مدل یادگیری مبتنی بر دفعات پاداش نبوده است. از آنجایی که مدل بهینه‌ی استفاده شده در این پژوهش مدل یادگیری مبتنی بر تعداد دفعات بوده و نه مدل یادگیری مبتنی بر مقدار پاداش که در پژوش‌های قبلی مبنای تحلیل بوده است، نتایج حاصل از این مطالعات را نمی‌توان به طور مستقیم با یکدیگر مقایسه کرد. بر این مبنا گرایش به سمت آسیب‌گریزی در مدل استفاده شده در این پژوهش معنای متفاوتی از گرایش به سمت آسیب‌گریزی در مدل یادگیری مبتنی بر مقدار، که در پژوهش‌های پیشین استفاده شده است، دارد. بنابراین، به منظور مقایسه‌ی دقیق‌تر باید بر داده‌های استفاده شده درکارهای پیشین نیز مدل یادگیری مبتنی بر دفعات را برازش داده و آن را با مدل یادگیری مبتنی بر مقدار مقایسه کرد. تنها در این صورت است که می‌توان مقایسه‌ی واقعی بین این پژوهش‌ها انجام داد.

علت دوم تفاوت نتایج در این مطالعه با مطالعات پیشین را می‌توان به درمانجو بودن معتادان در این مطالعه نسبت داد. معتادانی که در این پژوهش مورد ارزیابی شناختی قرار گرفتند افرادی بوده‌اند که به مرکز ملی مطالعات اعتیاد به منظور درمان مراجعه کرده‌اند. اساسا معتادانی که برای درمان مراجعه می‌کنند افرادی هستند که تحت تاثیر تجربه‌های گذشته‌ی خود در پی جلوگیری از آسیب‌های آینده-ی سوء مصرف مواد برآمده و برای درمان مراجعه کرده‌اند. بنابراین می‌توان رفتار آسیب‌گریزی مشاهده شده در آزمون را بر اساس میل به دوری از آسیب سوء مصرف مواد اعتیادآور که منجر به خصلت آسیب‌گریزی در این گروه شده است، توجیه کرد. در مطالعات پیشین، گروه معتاد از بین معتادان درمانجو نبوده و بنابراین می‌توان رفتاری متفاوت از معتادانِ درمانجویِ این پژوهش از آنها انتظار داشت.

علت کم بودن کارایی در گروه سالم در این پژوهش نسبت به گروه سالم در مطالعات غربی را می‌توان در تفاوت‌های بین فرهنگی جستجو کرد. رواجِ کم قمار در کشور ایران به علت اعتقادات دینی موجب کاهش آشناییِ عمومی با این مفهوم شده است. این امر به نوبه‌ی خود می‌تواند منجر به کم شدن کارایی افراد ایرانی در آزمون قمار شود.

از طرفی نسخه آزمون قمار استفاده‌شده در مرکز ملی مطالعات اعتیاد، نسبت به نسخه‌ای که در مطالعات پیشین انجام شده است، پیچیده‌تر است. بدین معنا که واریانس پاداش‌های دریافت شده در این آزمون نسبت به نسخه‌های غربی بیشتر بوده است. این امر به نوبه‌ی خود می‌تواند منجر به سخت-تر شدن یادگیری در این آزمون شود. همان‌طور که در بخش کارهای آینده ذکر خواهد شد، به منظور بررسی دقیق‌تر علت کم بودن کارایی، نیاز به طراحی نسخه‌ای از آزمون است که در آن توجه به دفعات پاداش و ضرر، به کارایی بالا در آزمون ختم شود.

۵-۵ خلاصه

در این فصل در ابتدا به تشریح مدل‌سازی شناختی پرداختیم. سپس داده‌های بدست آمده در آزمون قمار آیوا در معتادان و گروه شاهد را بیان کردیم. در ادامه به مدل‌سازی شناختی داده‌ها پرداخته و بر اساس مدل بهینه و مقادیر بهینه‌ی پارامترها، در مورد علت رفتار هر گروه توضیحاتی ارائه کردیم. در بخش پایانی نیز به بحث در مورد نتایج و مقایسه آن با دیگر مطالعه‌ها پرداختیم.

فصل ۶

نتیجه‌گیری و کارهای آینده

۶-۱ مقدمه

در این فصل از پایان‌نامه به تفکیک هر بخش به ارائه جمع‌بندی و ترسیم خطوطی برای ادامه‌ی پژوهش می‌پردازیم.

۶-۲ مدل‌سازی محاسباتی-عصبی اعتیاد به کوکائین

۶-۲-۱ خلاصه

با در نظر گرفتن سه فرض عصب‌شناسی:

۱. فعالیت فازیک نورون‌های دوپامینی کد کننده‌ی خطای پیش‌بینی هستند،
۲. مواد اعتیادآور باعث تولید مصنوعی دوپامین می‌شوند،
۳. مصرف مزمن مواد اعتیادآور باعث بالارفتن غیرعادی آستانه‌ی پاداش می‌شود،

مدلی برای اعتیاد به کوکائین پیشنهاد شد. مدل پیشنهاد شده علاوه بر انطباق با مصرف غیرحساس به هزینه در اعتیاد مزمن، کاهش ارزش پاداش‌های طبیعی و افزایش رفتار تکانشگرایانه را نیز توضیح می‌دهد. همچنین مدل پیش‌بینی می‌کند که پدیده‌ی بلوکه کردن در مورد پاداش مواد روی دهد.

مدل پیشنهاد شده بر خلاف مدل ردیش و گوتکین و همکاران اختلال در پردازش پاداش‌های طبیعی را ایجاب می‌کند. همچنین مقدار ارزش وضعیت مصرف مواد به طور نامحدود زیاد نمی‌شود، و همچون مدل گوتکین و همکاران پس از مصرف درازمدت، یادگیری به طور کلی از بین نمی‌رود.

۶-۲-۲ کارهای آینده

همان‌طور که در بخش ۴-۶ گفته شد، بررسی نحوه‌ی ارتباط بخش‌های دیگر مغز مانند بخش غشایی جلویی به مدل معرفی شده در این پژوهش، از قدم‌های مهم بعدی در رسیدن به مدلی واقعی‌تر از اعتیاد است. در این راستا در قدم اول باید مدل ارائه شده در [۸۰] (در بخش ۴-۶ مختصری درباره این مدل گفته شده است) به فرم یک یادگیری تفاضل زمانی بیزی بازنویسی شود که در نتیجه بتوان دخالت مواد در یادگیری را از طریق سیگنال خطا مدل‌سازی کرد. در قدم بعد مشخص شود که آیا دخالت مواد مقادیر قطعیت تخمین‌ها را غیر واقعی می‌کند یا خیر. پس از این مرحله می‌توان بر توانایی مدل ارائه شده در [۸۰] برای مدل‌سازی اعتیاد قضاوت کرد.

یکی از ایرادهای مهم که در مدل معرفی شده در این پژوهش و مدل‌های قبلی عصبی-محاسباتی است، عدم صحت سنجیِ کمّی این مدل‌هاست. بدین معنا که انطباق رفتار مدل با آزمایش‌های تجربی به طور کمّی بررسی نشده است. یکی از مهمترین منابع داده‌ای که می‌توان این گونه مدل‌ها را به وسیله آنها صحت‌سنجی کمّی کرد، چگونگی تغییر نرخ پاسخ در شرایط مختلف است. بدین معنا که با تغییر دوز مواد، سابقه‌ی مصرف پیشین، اندازه‌ی انتخاب دیگر رقابت‌کننده با مواد، نرخ پاسخ حیوان برای بدست آوردن مواد تغییر می‌کند. در حال حاضر تنها مدل موجود برای بررسی نرخ پاسخ در یادگیری تقویتی مدل معرفی شده توسط [۸۷] است. ولی متاسفانه در این مدل، انگیزه‌ی حیوان برای دریافت یک تقویت کننده در طول یک آزمون ثابت فرض شده است. این فرض به وضوح در مورد مواد که هر بار مصرف آن با سیری و در نتیجه کاهش انگیزه برای پاسخ همراه است، نقض می‌شود (برای توضیح مراجعه شود به ۳-۳). برای حل این مشکل می‌توان از مقایسه‌ی رفتار مدل در آزمایش‌هایی که پاسخ بلافاصله با پاداش همراه نیست، استفاده کرد. در این رده از آزمایش‌ها برنامه‌ریزی افزایشی و

مرحله‌ی دوم[۱۲۹] قرار دارد، که در آنها حیوان پس از یک پاسخ پاداش دریافت نمی‌کند. همچنین می‌توان از برنامه‌ریزی با نسبت ثابت (FR) که در آن هر تزریق با یک بازه‌ی زمانی که در آن حیوان قادر به تزریق نیست استفاده کرد. به هر حال اتخاذ هر رویکرد مستلزم ایجاد تغییراتی در مدل بیان‌شده در [۸۷] است که در کارهای بعدی می‌تواند مورد توجه قرار گیرد.

در مدل ارائه شده در این پژوهش و همین‌طور در دیگر مدل‌های عصبی-محاسباتی دوز مواد، متناظر با هیچ‌یک از متغیرهای مدل نیست. بدین معنا که مشخص نیست آیا با زیاد کردن دوز مواد مقدار D_{s_t} و یا r_t تحت تاثیر قرار می‌گیرد یا نه؟ این کاستی موجب شده است که نتوان مدل را بر اساس آزمایش‌هایی که رفتار حیوان را در دوزهای مختلف گزارش می‌کنند، صحت‌سنجی کرد. برای رفع این مشکل می‌توان به مدل‌سازی نقش دوز مواد پرداخت.

۶-۳ مدل‌سازی شناختی آزمون قمار آیوا

۶-۳-۱ خلاصه

در پژوهش انجام شده، از روش مدل‌سازی شناختی به منظور کشف علت ضعف عملکرد معتادان و نمونه‌های سالم ایرانی در آزمون قمار آیوا استفاده شد. نتایج حاصل از مدل‌سازی توانست برای رفتار مشاهده شده‌ی آزمون، توضیحی در سطح ساختار تصمیم‌گیری ارائه کند. بنابراین توضیح، هر دو گروه سالم و معتاد، برای ارزیابی گزینه‌های مختلف به مقدار پاداش و ضرر دریافت‌شده توجه نکرده و تنها از تعداد دفعات دریافت پاداش و ضرر به منظور تخمین ارزش هر عمل استفاده می‌کرده‌اند. علاوه بر این، در این پژوهش توضیحی به منظور علت عملکرد ضعیف‌تر معتادان نسبت به گروه سالم ارائه شده است. این کار بر اساس تحلیل حساسیت مدل صورت گرفته و نشان داد که علت ضعف عملکرد در معتادان گرایش غیرطبیعی به سمت آسیب‌گریزی در ارزیابی یک پاداش دریافت‌شده، بوده است.

[129] Second-order schedule

۶-۳-۲ کارهای آینده

در ادامه‌ی پژوهش ارائه شده می‌توان به تحلیل نظام تصمیم‌گیری معتادانی که درمانجو نیستند پرداخت. انجام این‌کار می‌تواند به روشن شدن علت تفاوت نتایج این پژوهش با نمونه‌های خارجی کمک کند.

به لحاظ دقت ریاضی، در این پژوهش میزان خوب بودن برازش مدل بهینه هر گروه به داده‌های آن گروه محاسبه نشده است. این کار به علت دشواری این امر در مدل‌های غیر خطی (مانند مدل‌های استفاده شده در این پژوهش) بوده است. در قدم‌های بعدی می‌توان به توسعه‌ی روابط ریاضی برای محاسبه میزان خوب بودن برازش توجه کرد.

به منظور تایید این فرض که علت کاستی در تصمیم‌گیری توجه به تعداد دفعات پاداش و نه مقدار آنها بوده است، لازم است که گونه‌ای از آزمون طراحی شود که در آن توجه به دفعات در تصمیم‌گیری به کارایی بالا در آزمون منجر شود. در این صورت اگر گروه سالم در آزمون کارایی بالایی داشته باشد، می‌توان نتیجه گرفت که برای آزمون‌دهندگان ایرانی، تعداد دفعات دریافت پاداش و ضرر معیار تصمیم‌گیری بوده است و کم بودن بودن امتیاز در این پژوهش به علت کم توجهی آزمودنی‌ها به آزمون و یا نا آشنایی با آزمون قمار نبوده است.

در پژوهش‌های بعدی به منظور تایید دستاوردهای مطرح شده در این پژوهش می‌توان از صحت-سنجی خارجی[۱۳۰] بهره جست. بدان معنا که مثلا ارتباط بین میزان انحراف به سمت آسیب‌گریزی و شاخص شدت اعتیاد[۱۳۱] بررسی شود؛ یا ارتباط میان IQ و میزان گرایش به سمت تصمیم‌گیری بر اساس تعداد دفعات مورد مطالعه قرار گیرد.

در نهایت نیز به منظور صحت‌سنجی مدل برازش‌یافته بر رفتار یک گروه می‌توان از ثبت فعالیت‌های مغزی در حین انجام آزمون استفاده کرد. این کار بدین صورت است که ارتباط میان سیگنال‌های درونی مدل و فعالیت بخش‌های مختلف مغز در حین آزمون بررسی می‌شود. در صورت کشف تناظر

[130] External validity
[131] Addiction Severity Index (ASI)

قابل توجه می‌توان نتیجه گرفت که محاسبه‌ی انجام شده توسط مدل به واقع در نظام عصبی پیاده-سازی می‌شود [۸۸].

مراجع

اختیاری، ح.، بهزادی، آ. جنتی، آ. مکری، آ. "دفعات باخت و مقادیر آن: کدام یک تأثیر منفی بیشتری بر ما می گذارد؟". فصلنامه تازه های علوم شناختی، شماره ۳ و ۴، ۱۷-۲۷، ۱۳۸۳.

[1] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR (Text Revision)*, American Psychiatric Publishing, Inc., 2000.

[2] E. Gwinnell and C. Adamec, *The Encyclopedia Of Addictions And Addictive Behaviors*, New York: Facts on File, 2006.

[3] J.J. Block, "Issues for DSM-V: Internet Addiction," *Am J Psychiatry*, vol. 165, Mar. 2008, pp. 306-307.

[4] R.M. Murray, P.D. Morrison, C. Henquet, and M. Di Forti, "Cannabis, the mind and society: the hash realities," *Nature Reviews. Neuroscience*, vol. 8, Nov. 2007, pp. 885-95.

[5] *The Social Impact of Drug Abuse*, Copenhagen: United Nations Office on Drugs and Crime (UNDCP), 1995.

[6] G. Bobashev, E. Costenbader, and B. Gutkin, "Comprehensive mathematical modeling in drug addiction sciences," *Drug and alcohol dependence*, vol. 89, 2007, pp. 102-6.

[7] S.H. Ahmed, G. Bobashev, and B.S. Gutkin, "The simulation of addiction: pharmacological and neurocomputational models of drug self-administration," *Drug and Alcohol Dependence*, vol. 90, Oct. 2007, pp. 304-11.

[8] P. Perez, A. Drey, A. Ritter, P. Dietze, and L.A. Mazerolle, *SimDrug: Exploring the Complexity of Heroin Use in Melbourne. Monograph 11*, Turning Point Drug and Alcohol Centre, 2005.

[9] P. Reuter, "The need for dynamic models of drug markets," *UNITED NATIONS PUBLICATION*, vol. LIII, 2001, pp. 1-10.

[10] J.P. Caulkins, D.A. Behrens, C. Knoll, G. Tragler, and D. Zuba, "Markov chain modeling of initiation and demand: the case of the U.S. cocaine epidemic," *Health care management science*, vol. 7, 2004, pp. 319-29.

[11] C. Rossi, "The role of dynamic modelling in drug abuse epidemiology," *United Nations Publications*, vol. LIV, 2002, pp. 33-44.

[12] C. Rossi, "Operational models for epidemics of problematic drug use :the Mover-Stayer approach to heterogeneity," *Socio-Economic Planning Sciences*, vol. 38, 2004, pp. 73-90.

[13] S.S. Everingham and C.P. Rydell, *Modeling the Demand for Cocaine*, Santa Monica: RAND Drug Policy Research Center, 1994.

[14] P.W. Kalivas and C. O'Brien, "Drug Addiction as a Pathology of Staged Neuroplasticity," *Neuropsychopharmacology*, vol. 33, 2007, p. 166—180.

[15] S.H. Ahmed, "NEUROSCIENCE: Addiction as Compulsive Reward Prediction," *Science*, vol. 306, 2004, p. 1901—1902.

[16] A. Rangel ,C. Camerer, and P.R. Montague, "A framework for studying the neurobiology of value-based decision making," *Nature Reviews. Neuroscience*, vol. 9, Jul. 2008, pp. 545-56.

[17] R.S. Sutton and A.G. Barto, *Reinforcement Learning: An Introduction*, Cambridge ,MA: MIT Press, 1998.

[18] J.E. Mazur, "Hyperbolic value addition and general models of animal choice," *Psychological Review*, vol. 108, Jan. 2001, pp. 96-112.

[19] S.P. Singh, "Reinforcement learning algorithms for average-payoff Markovian decision processes," *Proceedings of the twelfth national conference on Artificial intelligence (vol. 1)*, Seattle, Washington, United States: American Association for Artificial Intelligence, 1994, pp. 700-705.

[20] S. Mahadevan, "Average Reward Reinforcement Learning :Foundations, Algorithms, and Empirical Results," *Machine Learning*, vol. 22, Jan. 1996, pp. 159-195.

[21] R.S. Sutton, "Learning to predict by the methods of temporal differences," *Machine Learning*, vol. 3, 1988, pp. 9-44.

[22] C. Watkins, "Learning from Delayed Rewards," King's College, Cambridge, UK, 1989.

[23] I.P. Pavlov, *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*, Oxford University Press, 1927.

[24] L. Kamin, "Predictability, surprise, attention, and conditioning," *Punishment and aversive behavior*, B.A. Campbell and R.M. Church, eds., New York: Appleton-Century-Crofts, 1969, pp. 279-296.

[25] R.C. Rizley and R.A. Rescorla, "Associations in second-order conditioning and sensory preconditioning," *Journal of Comparative and Physiological Psychology*, vol. 81, Oct. 1972, pp. 1-11.

[26] R.A. Rescorla and A.R. Wagner, "A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement," *Classical*

Conditioning, 2: Current Research and Theory, A.H. Black and W.F. Prokasy, eds., New York: Appleton Century-Crofts, 1972, pp. 64–69.

[27] E.R. Kandel, J.H. Schwartz, and T.M. Jessell, *Principles of Neural Science*, Appleton & Lange, 1991.

[28] R.A. Wise, "Neuroleptics and operant behavior: the anhedonia hypothesis," *Behavioral and Brain Sciences*, vol. 5, 1982, pp. 39–87.

[29] K.C. Berridge and T.E. Robinson, "What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience?," *Brain Research. Brain Research Reviews*, vol. 28, Dec. 1998, pp. 309-69.

[30] W. Schultz, "Predictive reward signal of dopamine neurons," *Journal of Neurophysiology*, vol. 80, Jul. 1998, pp. 1-27.

[31] N.D. Daw, "Reinforcement learning models of the dopamine system and their behavioral implications," Carnegie Mellon University, 2003.

[32] P.R. Montague, P. Dayan, and T.J. Sejnowskw, "A framework for mesencephalic dopamine systems based on predictive Hebbian learning," *Journal of Neuroscience*, vol. 16, 1996, pp. 1936--1947.

[33] W. Schultz, P. Dayan, and P.R. Montague, "A Neural Substrate of Prediction and Reward," *Science*, vol. 275, 1997, pp. 1593-1599.

[34] N.D. Daw and D.S. Touretzky, "Long-term reward prediction in TD models of the dopamine system," *Neural Computation*, vol. 14 ,Nov. 2002, pp. 2567-83.

[35] Y. Niv and P.R. Montague, "Theoretical and empirical studies of learning," *Neuroeconomics: Decision Making and the Brain*, P.W. Glimcher, C. Camerer, P. Russell Alan, and E. Fehr, eds., Elsevier Science & Technology Books, 2008 ,pp. 249-329.

[36] S.M. McClure, N.D. Daw, and P.R. Montague, "A computational substrate for incentive salience," *Trends in Neurosciences*, vol. 26, Aug. 2003, pp. 423-8.

[37] A.D. Redish, "Addiction as a Computational Process Gone Awry," *Science*, vol ,2004 ,306 .pp. 1944-1947.

[38] B.S. Gutkin, S. Dehaene, and J. Changeux, "A neurocomputational hypothesis for nicotine addiction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, Jan. 2006, pp. 1106-11.

[39] E.J .Nestler, "Molecular basis of long-term plasticity underlying addiction," *Nature Reviews. Neuroscience*, vol. 2, Feb. 2001, pp. 119-28.

[40] R. Chen, M.R. Tilley, H. Wei, F. Zhou, F. Zhou, S. Ching, N. Quan, R.L. Stephens, E.R. Hill, T. Nottoli, D.D. Han ,and H.H. Gu, "Abolished cocaine reward in mice with a cocaine-insensitive dopamine transporter," *Proceedings of the National Academy of Sciences*, vol. 103, Jun. 2006, pp. 9333-9338.

[41] M.R. Tilley, B. O'Neill, D.D. Han, and H.H. Gu, "Cocaine does not produce reward in absence of dopamine transporter inhibition," *Neuroreport*, vol. 20, Jan. 2009, pp. 9-12.

[42] S. Stocker, "Cocaine's Pleasurable Effects May Involve Multiple Chemical Sites," *NIDA Notes*, vol. 14, 1999.

[43] N.D. Volkow, J.S. Fowler, G .Wang, J.M. Swanson, and F. Telang, "Dopamine in drug abuse and addiction: results of imaging studies and treatment implications," *Archives of Neurology*, vol. 64, Nov. 2007, pp. 1575-9.

[44] G.D. Stuber, R.M. Wightman, and R.M. Carelli, "Extinction of Cocaine Self-Administration Reveals Functionally and Temporally Distinct Dopaminergic Signals in the Nucleus Accumbens," *Neuron*, vol. 46, May. 2005, pp. 661-669.

[45] L.V. Panlilio, E.B. Thorndike, and C.W. Schindler, "Blocking of conditioning to a cocaine-paired stimulus: testing the hypothesis that cocaine perpetually produces a signal of larger-than-expected reward," *Pharmacology, Biochemistry, and Behavior*, vol. 86, Apr. 2007, pp. 774-7.

[46] Y. Mateo, C.M. Lack, D. Morgan, D.C.S. Roberts, and S.R. Jones, "Reduced dopamine terminal function and insensitivity to cocaine following cocaine binge self-administration and deprivation," *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, vol. 30, Aug. 2005, pp.63-1455 .

[47] N.D. Volkow, G.J. Wang, J.S. Fowler, J. Logan, S.J. Gatley, R. Hitzemann, A.D. Chen, S.L. Dewey, and N. Pappas, "Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects," *Nature*, vol. 386, Apr. 1997, pp. 830.3-

[48] Y. Shaham, U. Shalev, L. Lu, H. De Wit, and J. Stewart, "The reinstatement model of drug relapse: history, methodology and major findings," *Psychopharmacology*, vol. 168, Jul. 2003, pp. 3-20.

[49] A.D. Redish, S. Jensen, A. Johnson, and Z. Kurth-Nelson, "Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling," *Psychological Review*, vol. 114, Jul. 2007, pp. 784-805.

[50] J. Bergman and C.A. Paronis, "Measuring the reinforcing strength of abused drugs," *Molecular Interventions*, vol. 6, Oct. 2006, pp. 273-83.

[51] S.H. Ahmed and G.F. Koob, "Transition to drug addiction: a negative reinforcement model based on an allostatic decrease in reward function," *Psychopharmacology*, vol. 180, 2005, pp. 473-90.

[52] S.H. Ahmed and G.F. Koob, "Long-lasting increase in the set point for cocaine self-administration after escalation in rats," *Psychopharmacology*, vol. 146, Oct. 1999, pp. 303-12.

[53] S.H. Ahmed and G.F. Koob, "Transition from Moderate to Excessive Drug Intake: Change in Hedonic Set Point," *Science*, vol. 282, Oct. 1998, pp. 298-300.

[54] A.B. Norman and V.L. Tsibulsky, "The compulsion zone: a pharmacological theory of acquired cocaine self-administration," *Brain Research*, vol. 1116, Oct. 2006, pp. 143-52.

[55] H. Garavan, J. Pankiewicz, A. Bloom, J.K. Cho, L. Sperry, T.J. Ross, B.J. Salmeron, R. Risinger, D. Kelley, and E.A. Stein, "Cue-induced cocaine craving: neuroanatomical specificity for drug users and drug stimuli," *The American Journal of Psychiatry*, vol. 157, Nov. 2000, pp. 1789-98.

[56] R.Z. Goldstein, N. Alia-Klein, D. Tomasi, L. Zhang, L.A. Cottone, T. Maloney, F. Telang, E.C. Caparelli, L. Chang, T. Ernst, D. Samaras, N.K. Squires, and N.D. Volkow" ,Is decreased prefrontal cortical sensitivity to monetary reward associated with impaired motivation and self-control in cocaine addiction?," *The American Journal of Psychiatry*, vol. 164, Jan. 2007, pp. 43-51.

[57] G.F. Koob and M. Le Moal, "Plasticity of reward neurocircuitry and the 'dark side' of drug addiction," *Nature Neuroscience*, vol. 8, Nov. 2005, pp. 1442-4.

[58] R.L. Solomon, "The opponent-process theory of acquired motivation: the costs of pleasure and the benefits of pain," *The American Psychologist*, vol. 35, Aug. 1980, pp. 691-712.

[59] G.F. Koob and M. Le Moal, "Addiction and the Brain Antireward System," *Annual Review of Psychology*, vol. 59, 2008, pp. 29-53.

[60] N.D. Volkow, J.S. Fowler, G. Wang, and J.M. Swanson, "Dopamine in drug abuse and addiction: results from imaging studies and treatment implications," *Molecular Psychiatry*, vol. 9, Jun. 2004, pp. 557-69.

[61] M.A. Nader, D. Morgan, H.D. Gage, S.H. Nader, T.L. Calhoun, N. Buchheimer, R. Ehrenkaufer, and R.H. Mach, "PET imaging of dopamine D2 receptors during chronic cocaine self-administration in monkeys," *Nature Neuroscience*, vol. 9, Aug. 2006, pp. 1050-6.

[62] A.A. Grace, "The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving," *Addiction (Abingdon, England)*, vol. 95 Suppl 2, Aug. 2000, pp. S119-28.

[63] A.A. Grace, "The tonic/phasic model of dopamine system regulation: its relevance for understanding how stimulant abuse can alter basal ganglia function ",*Drug and Alcohol Dependence*, vol. 37, Feb. 1995, pp. 111-29.

[64] A.J. Smith, M. Li, S. Becker, and S. Kapur, "Linking Animal Models of Psychosis to Computational Models of Dopamine Function," *Neuropsychopharmacology*, vol. 32, May. 2006, pp. 54-66.

[65] W.J. Lynch and J.R. Taylor, "Decreased Motivation Following Cocaine Self-Administration Under Extended Access Conditions: Effects of Sex and Ovarian Hormones," *Neuropsychopharmacology*, vol. 30, Jan. 2005, pp. 927-935.

[66] A.M. Barr and A.G. Phillips" ,Withdrawal following repeated exposure to d-amphetamine decreases responding for a sucrose solution as measured by a progressive ratio schedule of reinforcement," *Psychopharmacology*, vol. 141, Jan. 1999, pp. 99-106.

[67] V. Deroche-Gamonet, D. Belin ,and P.V. Piazza, "Evidence for addiction-like behavior in the rat," *Science*, vol. 305, 2004, p. 1014—1017.

[68] J.R. Mantsch, A. Ho, S.D. Schlussman, and M.J. Kreek, "Predictable individual differences in the initiation of cocaine self-administration by rats under extended-access conditions are dose-dependent," *Psychopharmacology*, vol. 157, Aug. 2001, pp. 31-9.

[69] N.E. Paterson and A. Markou, "Increased motivation for self-administered cocaine after escalated cocaine intake," *Neuroreport*, vol. 14, Dec. 2003, pp. 2229-32.

[70] L.J.M.J. Vanderschuren and B.J. Everitt, "Drug Seeking Becomes Compulsive After Prolonged Cocaine Self-Administration," *Science*, vol. 305, 2004, pp. 1017-1019.

[71] J.H. Daruna and P.A. Barnes, "A neurodevelopmental view of impulsivity," *The Impulsive Client: Theory, Research, and Treatment*, W.G. McCown, J.L. Johnson, and M.B. Shure, eds., Washington, D.C.: American Psychological Association, 1993, p. 23.

[72] T.A. Paine, H.C. Dringenberg, and M.C. Olmstead, "Effects of chronic cocaine on impulsivity: relation to cortical serotonin mechanisms," *Behavioural Brain Research*, vol. 147, Dec. 2003, pp. 135-47.

[73] A. Logue, H. Tobin, J. Chelonis, R. Wang, N. Geary, and S. Schachter, "Cocaine decreases self-control in rats: a preliminary report," *Psychopharmacology*, vol. 109, Oct. 1992, pp. 245-247.

[74] N.W. Simon, I.A. Mendez, and B. Setlow, "Cocaine exposure causes long-term increases in impulsive choice," *Behavioral neuroscience*, vol. 121, Jun. 2007, pp. 543–549.

[75] D. Belin, A.C. Mar, J.W. Dalley, T.W. Robbins, and B.J. Everitt, "High impulsivity predicts the switch to compulsive cocaine-taking," *Science (New York, N.Y.)*, vol. 320, Jun. 2008, pp. 1352-5.

[76] T.K. Das, A. Gosavi, S. Mahadevan, and N. Marchalleck, "Solving semi-markov decision problems using average reward reinforcement learning," *Management Science*, vol. 45, 1999, pp. 560--574.

[77] J.W. Dalley, T.D. Fryer, L. Brichard, E.S.J. Robinson, D.E.H. Theobald, K. Lääne, Y. Peña, E.R. Murphy, Y. Shah, K. Probst ,I. Abakumova, F.I. Aigbirhio, H.K. Richards, Y. Hong, J. Baron, B.J. Everitt, and T.W. Robbins, "Nucleus accumbens D2/3 receptors predict trait impulsivity and cocaine reinforcement," *Science (New York, N.Y.)*, vol. 315, Mar. 2007, pp. 1267-70.

[78] N.D .Daw, S. Kakade, and P. Dayan, "Opponent interactions between serotonin and dopamine," *Neural Networks: The Official Journal of the International Neural Network Society*, vol. 15, 2002, pp. 603-16.

[79] Y. Pelloux, B.J. Everitt, and A. Dickinson, "Compulsive drug seeking by rats under punishment: effects of drug taking history," *Psychopharmacology*, vol. 194, 2007, pp. 127-37.

[80] N.D. Daw, Y. Niv, and P. Dayan, "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control," *Nature Neuroscience*, vol. 8, Dec. 2005, pp. 1704-11.

[81] A. Bechara, A.R. Damasio, H. Damasio, and S.W. Anderson, "Insensitivity to future consequences following damage to human prefrontal cortex.," *Cognition*, vol. 50, 1994, pp. 7-15.

[82] J.R. Busemeyer and J.C. Stout, "A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task," *Psychological assessment*, vol. 14, 2002, pp. 253-62.

[83] K. Kalidindi and H. Bowman, "Using e-greedy reinforcement learning methods to further understand ventromedial prefrontal patients' deficits on the Iowa Gambling Task," *Neural Netw.*, vol. 20, 2007, pp. 676-689.

[84] A. Bechara, S. Dolan, N. Denburg, A. Hindes, S.W. Anderson, and P.E. Nathan" ,Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers," *Neuropsychologia*, vol. 39, 2001, pp. 376-89.

[85] J.C. Stout, S.L. Rock, M.C. Campbell, J.R. Busemeyer, and P.R. Finn, "Psychological processes underlying risky decisions in drug abusers," *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, vol. 19, Jun. 2005, pp. 148-57.

[86] J.C. Stout, J.R. Busemeyer, A. Lin, S.J. Grant, and K.R .Bonson, "Cognitive modeling analysis of decision-making processes in cocaine abusers," *Psychonomic bulletin & review*, vol. 11, 2004, pp. 742-7.

[87] Y. Niv, N.D. Daw, D. Joel, and P. Dayan, "Tonic dopamine: opportunity costs and the control of response vigor.," *Psychopharmacology (Berl)*, vol. 191, 2007, p. 507—520.

[88] G. Corrado and K. Doya, "Understanding neural coding through the model-based analysis of decision making," *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, vol. 27, Aug. 2007, pp. 8178-80.

[89]C.P. O'Brien, N. Volkow, and T. Li, “What's in a word? Addiction versus dependence in DSM-V,” *The American Journal of Psychiatry*, vol. 163, May. 2006, pp. 764-5.

پیوست الف- مدل‌های یادگیری

مدل‌های یادگیری استفاده‌شده در ارزیابی شناختی

به لحاظ بیولوژیکی و روانی، ساختارهای مختلفی در مغز وظیفه‌ی بازنمایی مقادیر مثبت و منفی را به عهده دارند که این پدیده باعث وزن‌دهی متفاوت مقادیر مثبت و منفی می‌شود:

$$r_t(a) = w.r_t^+(a) + (1-w).r_t^-(a)$$

$r_t^+(a)$ و $r_t^-(a)$ به ترتیب، پاداش و تنبیه دریافت‌شده توسط عامل، پس از انجام عمل a می‌باشند. در صورتی که پارامتر w ($0 \leq w \leq 1$) که به آن وزن ارزش‌گذاری می‌گوییم، مقداری نزدیک به یک داشته باشد، نشان‌دهندهٔ رفتار پاداش‌جویی و در صورتی که مقداری نزدیک به صفر داشته باشد، نشان‌دهندهٔ رفتار آسیب‌گریزی عامل می‌باشد. یادگیری بر اساس سیگنال نهایی ارزش ($r_t(a)$)، صورت می‌گیرد. هدف اصلی از یادگیری تخمین ارزش هر دسته از کارت است که این تخمین با $Q(a)$ نشان داده می‌شود (a می‌تواند هر یک از دسته کارت‌ها باشد). در ادامه مدل‌های یادگیری شرح داده شده‌اند.

۱- **میانگین‌گیری ساده:** بر اساس این روش، ارزش هر عمل به‌صورت میانگین تمام سیگنال‌های ارزش تجربه‌شده توسط عامل، تخمین زده می‌شود:

$$Q_T(a) = \frac{1}{K_a}\sum_{t=1}^{T} r_t(a)$$

K_a تعداد دفعاتی است که عمل a تا پیش از زمان T توسط عامل تجربه شده است.

۲- **یادگیری مبتنی بر واریانس**: این روش، مدل‌سازی رفتار مخاطره‌جویانه را هدف قرار داده است و لذا با افزایش واریانس پاداش‌های دریافت‌شده به ازای یک عمل، ارزش آن عمل نزد عامل بیشتر می‌شود:

$$Q_T(a) = \frac{1}{K_a}\sum_{t=1}^{T}\left(r_t(a) - \bar{r}_t(a)\right)^2$$

$\bar{r}_t(a)$ بیان‌کننده میانگین نتایج اخذ شده در اثر انجام عمل a می‌باشد.

۳- **یادگیری مبتنی بر فرکانس:** در این روش، به جای میزان خطا از تعداد دفعاتی که در اثر انجام یک عمل، پاداش مثبت و یا منفی گرفته شده است، به منظور ارزیابی ارزش یک عمل استفاده می‌شود:

$$Q_t(a) = \begin{cases} Q_{t-1}(a) + 1 & r_t(a) > 0 \\ Q_{t-1}(a) - 1 & r_t(a) < 0 \\ Q_t(a) & else \end{cases}$$

۴- **یادگیری مبتنی بر سیگنال خطا:** در این روش، یادگیری در جهت تصحیح خطایی که در پیش‌بینی ارزش یک عمل اتفاق افتاده است، صورت می‌گیرد:

$$Q_t(a) = Q_{t-1}(a) + \gamma(r_t(a) - Q_{t-1}(a))$$

که در رابطه‌ی فوق γ نرخ یادگیری است.

۵- **یادگیری مبتنی بر فرکانس خطا:** این روش، همانند مدل مبتنی بر سیگنال خطا است؛ لیکن از تعداد دفعات دریافت پاداش و یا تنبیه برای به‌روزرسانی ارزش عمل استفاده می‌شود:

$$Q_t(a) = \begin{cases} Q_{t-1}(a) + \gamma(1 - Q_{t-1}(a)) & r_t(a) > 0 \\ Q_{t-1}(a) - \gamma(1 + Q_{t-1}(a)) & r_t(a) < 0 \\ (1-\gamma)Q_{t-1}(a) & else \end{cases}$$

۶- **یادگیری معکوس:** درصورتی که ارزش تخمین‌زده شده برای یک عمل، در جهت مخالف ارزش تحقق یافته‌ی آن باشد، سرعت یادگیری عامل کاهش خواهد یافت:

$$\begin{aligned} &\text{if} && sign(Q_{t-1}(a)) = sign(Z) \quad then \\ &&& Q_t(a) = Z \\ &\text{else} \\ &&& Q_t(a) = Q_{t-1}(a) + \lambda.\gamma(r_t(a) - Q_{t-1}(a)) \\ &\text{where} \\ &&& Z = Q_{t-1}(a) + \gamma(r_t(a) - Q_{t-1}(a)) \end{aligned}$$

روش تخمین پارامتر

با فرض اینکه فرآیند تصمیم‌گیری هر عامل انسانی، مبتنی بر یکی از روش‌های یادگیری و یکی از روش‌های انتخاب عمل است، تعدادی پارامتر مجهول باقی می‌ماند. بایستی پارامترهای مجهول به نحوی مقداردهی شوند که بیشترین شباهت میان رفتار عامل انسانی –در آزمون قمار- و رفتار مدل یادگیری تقویتی وجود داشته باشد. بدین منظور، روش حداکثر درست‌نمایی[1] بکارگرفته شده است. به منظور محاسبه‌ی تابع درست‌نمایی، لازم است خروجی سیستم‌ها (برای دو گروه سالم و معتاد) در هر مقطع از زمان (T=۱۰۰) و نیز تابع توزیع احتمال خروجی مدل به ازای هر بردار پارامترها (P_i). در هر مقطع از زمان دانسته شوند. خروجی سیستم‌ها که با نماد $O_{C,i,t}(a_k)$ برای گروه سالم و نماد $O_{A,i,t}(a_k)$ برای گروه معتاد نشان داده شده است، برابر با کل تعداد دفعاتی که عمل a_k در زمان t توسط هر یک از دو گروه انتخاب شده است. همچنین جهت تخمین تابع توزیع احتمال مدل که آن را با $Pr_t(P_j, a_k)$ نشان می‌دهیم، بنا به ماهیت تصادفی مدل، به ازای هر بردار پارامتر P_i، مدل ۳۰۰۰ بار اجرا شده است و از میانگین خروجی‌های حاصل‌شده، به عنوان تخمین $Pr_t(P_j, a_k)$ استفاده شده است. تابع درست‌نمایی بصورت زیر معرفی می‌شود:

$$f^{Control}(y \mid P_j) = \prod_{i=1}^{N} \prod_{t=1}^{100} \prod_{k=1}^{4} Pr_t(P_j,\ a_k)^{O_{C,i,t}(a_k)}$$

$$f^{SDI}(y \mid P_j) = \prod_{i=1}^{M} \prod_{t=1}^{100} \prod_{k=1}^{4} Pr_t(P_j,\ a_k)^{O_{A,i,t}(a_k)}$$

که در آن، k ($1 \le k \le 4$)، اندیس مربوط به انتخاب یکی از چهار عمل *A*، *B*، *C* و *D* توسط عامل می‌باشد. N و M به ترتیب تعداد افراد گروه‌های سالم و معتاد می باشند. بر اساس قانون حداکثرسازی درست‌نمایی خواهیم داشت:

$$P^*_{Control} = \arg\max_j f^{Control}(y|P_j)$$

$$P^*_{SDI} = \arg\max_j f^{SDI}(y|P_j)$$

[1] Maximum likelihood

جهت تخمین پارامترهای مدل‌هایی که کمتر از چهار درجهٔ آزادی می‌باشند، از روش جستجوی کامل فضا و برای مدل‌های دارای درجهٔ آزادی چهار و بیشتر از آن، به دلیل پیچیدگی محاسباتی بیش از حد روش جستجوی جامع، از الگوریتم ژنتیک استفاده شده است.

به منظور انتخاب بهترین مدل برای توصیف هر دسته از عامل‌ها، معیار اطلاعاتی بیزی که علاوه بر درست‌نمایی، پیچیدگی مدل را نیز لحاظ می‌کند، استفاده شده است:

$$BIC = -2\ln f\left(y \mid \hat{P}_j\right) + k\ln n$$

در رابطهٔ فوق، n تعداد داده‌ها و k درجهٔ آزادی مدل -تعداد پارامترها- می‌باشند.

پیوست ب- مقاله‌های مستخرج از پایان‌نامه

مقاله‌ی زیر بر اساس فصل ۵ بوده و در کنگره علوم شناخی ۲۰۰۸ (CogSci 2008) به صورت پوستر ارائه شده است.

Proceedings of the 30th Annual Conference of the Cognitive Science Society (p. 1094). Austin, TX: Cognitive Science Society.

Understanding Addictive Behavior on the Iowa Gambling Task Using Reinforcement Learning Framework

Amir Dezfouli[1], Mohammad Mahdi Keramati[2], Hamed Ekhtiari[3], Hooman Safaei[3], Caro Lucas[1]

[1]Center of Excellence for Control and Intelligent Processing,
Department of Electrical and Computer Engineering, University of Tehran, Iran

[2]School of Management and Economic, Sharif University of Technology, Iran
[3]Cognitive Assessment Laboratory, Iranian National Center for Addiction Studies, Iran

Abstract- Neurocognitive decision-making disorders in Iowa Gambling Task (IGT) can be better understood in the light of computational modeling methods. In this study, we use Reinforcement Learning (RL) framework to decompose subjects' behavior into its underlying factors. Both healthy subjects and Substance Dependent Individuals (SDIs) show poor performance in the task, with significant decline in SDIs (net score = -2.3) compared with control subjects (net score = 6.2). Fitting various models of RL family, the results show that for both groups, frequency-based learning model coupled with softmax exploration strategy for action selection is the best descriptor model for choices of subjects in the task (based on Bayesian Information Criterion). So, being under the influence of reinforcer frequency instead of its magnitude is the major factor behind poor performance of subjects. In addition, sensitivity analysis shows that the performance of the best fitted model is sensitive to the valence weight parameter in SDIs. The estimated value of the parameter reveals that higher deviation of SDIs to harm-avoidance characteristic (in relation to healthy subjects) causes performance difference between two groups. Neural and cultural discussions are also presented to explain the results.

چکیده‌ی زیر در کنفرانس عصب‌شناسی ۲۰۰۸ (Neuroscience 2008) به صورت پوستر ارائه شده است.

SfN: Neuroscience 2008. Washington DC November, 2008

A neurocomputational model for decreased harm avoidance in addicts

Amir Dezfouli[1], Payam Piray[1], Mohammad Mehdi Kramati[2], Hamed Ekhtiari[3], Caro Lucas[1], Azarakhsh Mokri[4,5]

[1]Center of Excellence for Control and Intelligent Processing,
Department of Electrical and Computer Engineering, University of Tehran, Iran

[2]School of Management and Economic, Sharif University of Technology, Iran
[3]Cognitive Assessment Laboratory, Iranian National Center for Addiction Studies, Iran
[4]Department of Psychiatry, Tehran University of Medical Sciences
[5]Department of Clinical Sciences, Iranian National Center for addiction Studies

Abstract- Lack of harm avoidance is one of the most important decision making deficits in addicts. In previous computational models of addiction, this property is explained by high incentive value of drug of abuse which grows unbounded with consumption and compensates punishment of harmful consequences. But this theory is not confirmed by some animal models of drug self-administration which report unchanged incentive value of the drug before and after subjects show compulsive drug seeking. In this study we are to propose a computational model based on theoretical models of dopamine system and Bayesian Q-learning to shed light on the basis of this phenomenon.

The agent learns values of actions in Bayesian fashion. Its action selection is guided by both expected state-action values and expected gain from exploring an action. The later indicates that the agent takes an action with low expected value if it predicts that the lost utility can be compensated by future gains of obtained information after executing the action (in the form of action selection policy improvement).

We modeled rewarding pattern of natural rewards and cocaine in the manner that they have the same maximum reward value. Drug induced transient increase in dopamine was modeled by impulse rewards those vanish as the expected value saturates at its maximum level. The agent was simulated under situation where it should choose between avoiding harm(freezing) and pressing seeking lever which causes receiving a punishment followed by reward delivery. The results show in the case of natural reward, the agent chooses freezing, but in the case of drug, although expected value of pressing seeking lever dose not differ form natural rewards, the agent chooses drug seeking. Tracking agent's action selection policy revealed that such risk-taking behavior is because of high predicted gain of policy improvement after experiencing cocaine intake. Roughly speaking, the addicts chooses drug because it unrealistically predicts that after consumption, estimated value of drug taking will be updated to the value higher than freezing. It means that it hopes the punishment to be removed in the future. So, it keeps drug taking against harmful consequences with the hope of gaining utility more than harm avoidance. Our model proposes a biologically-inspired computational base for compulsive drug seeking and risky decision-making in addicts.

چکیده‌ی زیر در سومین کنفرانس علوم شناختی (ICCS) ارائه شده است.

The Third International Conference of Cognitive Science, Tehran, Iran

Computational modeling of cocaine addiction using reinforcement learning framework

Amir Dezfouli[1], Payam Piray[1], Mohammad Mehdi Kramati[2], Hamed Ekhtiari[3], Caro Lucas[1], Azarakhsh Mokri[4,5]

[1]Center of Excellence for Control and Intelligent Processing,
Department of Electrical and Computer Engineering, University of Tehran, Iran

[2]School of Management and Economic, Sharif University of Technology, Iran
[3]Cognitive Assessment Laboratory, Iranian National Center for Addiction Studies, Iran
[4]Department of Psychiatry, Tehran University of Medical Sciences
[5]Department of Clinical Sciences, Iranian National Center for addiction Studies

Abstact- ***Objective:*** under two assumptions, first, phasic activity of dopaminergic neurons in ventral tegmental area (VTA) qualitatively corresponds to error signal employed in value learning process and second, drug consumption leads to increase of dopamine in VTA, we propose a neurocomputational model for drug addiction. ***Method:*** temporal difference reinforcement learning (TDRL) framework was used. Drug induces changes was modeled by adding an uncompensatable parameter to the error signal term in TDRL. Also, the level against which rewards are compared was introduced into TDRL using an additional term. ***Results:*** simulations show that the behavior of the model is satisfactorily compatible with the animal models of drug self-administration, especially compulsive drug seeking. Some other aspects of addition such as down-regulation of reward system are also addressed by the model. ***Conclusion:*** many decision-making deficits of drug addiction can be explained based on the two mentioned assumptions using computational modeling approach. Also, the model presents explicit behavioral predictions which can be tested both on human and animals.

چکیده- هدف: براساس دو فرض، اول آنکه فعالیت‌های نرون‌های دوپامینی درناحیه تگمنتوم شکمی (VTA) کد کنندهٔ سیگنال خطا در فرایند یادگیری هستند، و دوم آنکه مصرف کوکائین باعث زیاد شدن مقدار دوپامین در این بخش می‌شود، در این پژوهش یک مدل عصبی-محاسباتی برای اعتیاد به کوکایین ارائه شده است. **روش**: به منظور مدل‌سازی فرایندهای مغزی، از مدل یادگیری تقویتی تفاضل زمانی (temporal difference reinforcement learning) استفاده شده است. تاثیرات شیمیایی کوکائین با استفاده از یک فاکتور جبران ناپذیر در سیگنال خطا بازنمایی محاسباتی شده است. به علاوه، سطحی که پاداش‌ها با آن مقایسه می‌شود، توسط یک پارامتر اضافی به الگوی یادگیری تقویتی تفاضل زمانی اضافه شده است. **نتایج**: شبیه‌سازی مدل ارائه شده نشان می‌دهد که رفتار آن انطباق خوبی با مدل‌های حیوانی خودتزریقی دارو دارد. به علاوه، کاهش حساسیت در مسیر پاداش و برخی ویژگی‌های دیگر اعتیاد به مواد نیز توسط مدل ارائه شده، توضیح داده می‌شود. **نتیجه‌گیری**: با تکیه بر دو فرض بیان شده می‌توان توسط رویکرد مدل‌سازی محاسباتی، ویژگی‌های اصلی اعتیاد به کوکائین را توضیح داد. به علاوه مدل ارائه شده، پیش-بینی های صریحی را بیان می‌دارد که هم در حیوانات و هم در انسان قابل صحت‌سنجی می‌باشند.

مقاله‌ی زیر در مجله Neural Computation چاپ خواهد شد.

To Appear in Neural Computation

A Neurocomputational Model for Cocaine Addiction

Amir Dezfouli[1], Payam Piray[1], Mohammad Mehdi Kramati[2], Hamed Ekhtiari[3], Caro Lucas[1], Azarakhsh Mokri[4,5]

[1]Center of Excellence for Control and Intelligent Processing,
Department of Electrical and Computer Engineering, University of Tehran, Iran

[2]School of Management and Economic, Sharif University of Technology, Iran
[3]Cognitive Assessment Laboratory, Iranian National Center for Addiction Studies, Iran
[4]Department of Psychiatry, Tehran University of Medical Sciences
[5]Department of Clinical Sciences, Iranian National Center for addiction Studies

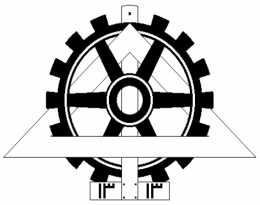
Abstract- Based on the dopamine hypotheses of cocaine addiction and the assumption of decrement of brain reward system sensitivity after long-term drug exposure we propose a computational model for cocaine addiction. Utilizing average reward temporal difference reinforcement learning, we incorporate the elevation of basal reward threshold after long-term drug exposure into the previous model of drug addiction proposed by Redish. Our model is consistent with the animal models of drug seeking under punishment. In the case of non-drug reward, the model explains increased impulsivity after long-term drug exposure. Furthermore, the existence of blocking effect for cocaine is predicted by our model.

Abstract

Addiction can be characterized with the compulsive drug seeking and taking behavior. A theory of addiction provides the reductive links across addictive behavior and structural evidence at neural or psychological level.

This thesis includes two sections. In the first section, based on the dopamine hypotheses of cocaine addiction and the assumption of decrement of brain reward system sensitivity after long-term drug exposure we propose a computational model for cocaine addiction. Utilizing average reward temporal difference reinforcement learning, we incorporate the elevation of basal reward threshold after long-term drug exposure into the previous model of drug addiction proposed by Redish. Our model is consistent with the animal models of drug seeking under punishment. In the case of non-drug reward, the model explains increased impulsivity after long-term drug exposure. Furthermore, the existence of blocking effect for cocaine is predicted by our model.

In the second section, we investigated behavior of addicts in Iowa Gambling Task. We used Reinforcement Learning (RL) framework to decompose subjects' behavior into its underlying factors. Both healthy subjects and Substance Dependent Individuals (SDIs) show poor performance in the task, with significant decline in SDIs (net score = -2.3) compared with control subjects (net score = 6.2). Fitting various models of RL family, the results show that for both groups, frequency-based learning model coupled with softmax exploration strategy for action selection is the best descriptor model for choices of subjects in the task (based on Bayesian Information Criterion). So, being under the influence of reinforcer frequency instead of its magnitude is the major factor behind poor performance of subjects. In addition, sensitivity analysis shows that the performance of the best fitted model is sensitive to the valence weight parameter in SDIs. The estimated value of the parameter reveals that higher deviation of SDIs to harm-avoidance characteristic (in relation to healthy subjects) causes performance difference between two groups.



University of Tehran
College of Engineering
School of Electrical and
Computer Engineering

Computational Modeling of Drug Abuse

By:

Amir Dezfouli

Under Supervision of:

Dr. Caro Lucas

Co-advisor:

Dr. Azarakhsh Mokri

Dissertation submitted to the Graduate Studies Office
in partial fulfillment of the requirements for the degree of
Master of Science in Computer Engineering,
Artificial Intelligence and Robotics Branch

May, 2009