Towards human-like artificial intelligence using StarCraft 2

Henrik Siljebråt
Goldsmiths, University of London
London, United Kingdom

Caspar Addyman
Goldsmiths, University of London
London, United Kingdom

Alan Pickering
Goldsmiths, University of London
London, United Kingdom

ABSTRACT

On our path towards artificial general intelligence, video games have become excellent tools for research. Reinforcement learning (RL) algorithms are particularly successful in this domain, with the added benefit of having fairly well established biological foundations. To improve how artificial intelligence research and the cognitive sciences can inform each other, we argue the StarCraft II Learning Environment is an ideal candidate for an environment where humans and artificial agents can be tested on the same tasks. We present an upcoming study using this environment, where the goal is to investigate how RL can be extended to enable abstract human abilities such as moments of insight. We claim this is valuable for advancing our understanding of both artificial and natural intelligence, thereby leading to improved models of player behaviour and for general video game playing.

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**; • **Applied computing → Psychology**;

KEYWORDS

artificial intelligence, StarCraft, reinforcement learning

ACM Reference Format:
Henrik Siljebråt, Caspar Addyman, and Alan Pickering. 2018. Towards human-like artificial intelligence using StarCraft 2. In *Foundations of Digital Games 2018 (FDG18), August 7–10, 2018, Malmö, Sweden.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3235765.3235811

1 INTRODUCTION

How do humans learn to play video games? The question may seem insignificant but holds the key to advances in our understanding of both artificial and natural intelligence. Video games are excellent test beds for artificial intelligence (AI; [17]) as evidenced by systems reaching human-level performance on Atari [23]. However, there are still many things these systems cannot do [18] and with video games specifically, we face challenges such as achieving human-like learning rate and general video game playing (GVGP; [20]). Since humans can not only learn to play different types of games but also solve many other problems, they (we) are the best model of a general problem solver available. From the breadth of challenges games pose we can use artificial systems based on cognitive architectures for GVGP as well as artificial general intelligence (AGI) [30].

In order to use humans properly as models for GVGP agents there are three things we need to be clear on; first we have to use a human-centric approach, meaning that what we model is a human playing a game which in turn means the interface needs to be as human-like as possible. Second, we are interested in biologically constrained systems based on findings in the cognitive sciences (from hereon used as collective term for neuroscience, psychology, behavioural economics etc.). This way, GVGP can test predictions made in the theories from cognitive science so the two fields can inform each other [14]. Third, to keep the system as general as possible, we need as little pre-programmed knowledge as possible so the game can be learned and played in a so called end-to-end fashion (usually implying the input is the game screen pixels and output is motor commands to a game controller), but see the discussion section for an alternate interpretation.

By now, reinforcement learning (RL; [35]) is fairly well established as an algorithm used in the brain's dopaminergic projection systems to allow for learning by reward, generally called the reward prediction error (RPE) hypothesis [25, 31]. RL has further been used successfully in many practical applications of game playing agents [23**?**] enabling them to learn to play end-to-end. It is therefore an excellent candidate algorithm to use as a basic building block for an artificial system. While RL works well for finding the value of actions depending on the state, it is less clear how RL interacts with other brain systems to find higher order task structures [41]. We can frame this in the form of Kahneman's [16] dual systems theory where RL is System 1 (fast but habitual) and we are aiming to find System 2 (slow but flexible).

This paper will explain what challenges real time strategy (RTS) games pose and how previous research on *StarCraft* is difficult to compare with human experimental data. We argue the newly released StarCraft 2 Learning Environment (SC2LE; [40]) is a better framework for such research and present the design of an upcoming experiment using *StarCraft 2* and human participants playing a custom map. The results of this experiment may enlighten how RL can be enhanced by additional systems on the path towards GVGP and AGI.

2 PREVIOUS RESEARCH WITH STARCRAFT

Strategy games are rich problem domains because they involve decision making on different time scales; players need to consider both their long-term strategy and their actions in the short term [4]. Games such as *Chess* are turn based - players take turns performing actions - whereas *StarCraft: Brood War* (SCBW; [2]) and its sequel *StarCraft 2* (SC2; [3]) are real time strategy (RTS) games where the environment and opponents do not wait for the player. The versions of *StarCraft* are similar enough that general concepts apply to both, so unless otherwise specified, the name *StarCraft* implies both. Furthermore, since the main *StarCraft* game type in tournaments is one versus one play, that is what we focus on here.

FDG'18, August 7-10, 2018, Malmö, Sweden

ACM ISBN 978-1-4503-6571-0/18/08.
https://doi.org/10.1145/3235765.3235811

Starcraft is set in a science fiction universe where three factions - Terran, Zerg and Protoss - compete for dominance of the galaxy. Players see the game from an isometric perspective and start in different locations of a map with various terrain features like hills, valleys and islands. To win the player must gather resources, construct buildings, research upgrades, train an army and eliminate all enemy buildings. Most games end earlier, as players commonly forfeit when they think the game is lost. *StarCraft* can thus be seen as a more intense form of *Chess*, as both involve interactions between long term strategy and short term decisions such as sacrificing pieces in *Chess*. The *StarCraft* games have millions of players, including professionals competing in tournaments for hundreds of thousands of US dollars. This means we can compare artificial agents to players of many different skill levels and also have access to a wealth of player data in the form of match recordings (replays).

There are several challenges in *StarCraft* for humans and artificial agents alike. First, the observation and action spaces are enormous; orders of magnitude larger than in *Go* [26, 28, 40]. Second, the real time nature of the games significantly shortens the time allowance for deliberation. Third, players can only see part of the map where the camera is (but do have an overview in the form of a low resolution minimap) and visibility is also occluded by "fog of war" except for any region where the player has units. This partial observability and imperfect information means players have to explore and scout their opponent to react appropriately. Fourth, games last anywhere from a few minutes to around 40 minutes meaning thousands of image frames and action sequences, and early actions can have long term consequences, posing a challenge for credit assignment [40].

Because of these challenges, most previous AI research has focused on solving subparts of the game such as controlling units in combat situations (shorter term tactical decisions) or learning build orders (game plans, longer term strategic decisions) from replays [26, 28, 40]. Though many valuable findings have been made, most methods involve implementing human expert knowledge. There are several examples of RL use, but these are either tailored to controlling units [24, 32, 42] or frame the problem as one where individual units are agents instead of modelling the player [11, 27, 39].

Bots (agents that can play the full game) for SCBW use combinations of different approaches for strategic and tactical decisions, usually hard-coded knowledge amplified by algorithms [5, 28]. This often leads to integration issues for when the tactical and strategic systems have to interact [28]. There are examples of success such as LetaBot[1] which uses Monte Carlo Tree Search to plan squad movement, and text mining to extract build orders.

None of the mentioned approaches have any claims of their systems being biologically plausible, which is what we are interested in here. It should be noted that watching replays is common among human players both for entertainment and training purposes. So the fact that bots learn strategies and build orders from replays, can be seen as fairly "human-like" but still suffer from interface issues, as will be explained in the next section below. There is at least one reported attempt of using the cognitive architecture Soar [44] for OpenRTS (but see Soar-SC[2]). However, such holistic models of the human brain are not appropriate for our purposes as our aim is to build from the ground up.

3 SC2LE PROVIDES THE HUMAN TOUCH

For our purposes, there are two main issues with much of the previous research. First, as [28] point out; not many techniques can play the full game and most forms of partial systems were evaluated with custom metrics making it difficult to compare results. Complete systems have for years been able to compete in competitions such as CIG, AIIDE and SSCAIT (see [6] for an overview), where submitted bots play hundreds or thousands of games against other entries and the highest win ratio takes the crown. This is an equal opportunity situation, but a slow feedback loop [28]. SSCAIT does run all year, but one would still need to wait until a bot has played many games to evaluate it. There are attempts towards standardization like [38] but we then run into the second issue of other environments like BWAPI[3] and Torchcraft [36] not providing observations and action inputs in an interface similar to the one humans use. As there is no player camera, all units and structures are available without having to move the view to the relevant part of the map. This means it is difficult to compare human experimnetal data with artificial agents.

The StarCraft II Learning Environment (SC2LE; [40]) was developed to provide a more "human-like" environment for an artificial agent to learn in. It simplifies observations by providing abstracted RGB images called feature layers, divided in two main categories for the minimap and game screen, each representing aspects such as unit type, owner and visibility. These top down orthographic projections maintain the spatial and graphical aspects of the true game screen frames and are delivered as NxM pixel matrices, meaning they do not exactly match what humans see. However, with N and M greater than 64 the feature layers have sufficient resolution for a human to play using this interface [40], meaning we can still call it "human-like". The action space is defined by circa 300 actions identified from thousands of replays and provided with each observation of the game state, to model the behaviour of the human interface where available actions are context specific depending on what unit or building is selected, if any. Rewards are given either after the game as win/tie/loss or based on an in-game score, the "Blizzard score", which human players can see after the game has ended but is provided for agents during the game for a less sparse reward structure. The Blizzard score increases with mined resources and decreases when losing units and/or buildings.

Using the easiest built-in AI of SC2 in a Terran vs Terran matchup, none of the artificial neural network architectures trained with RL used by [40] learned to win, but amusingly the best result was one that learned to fly off with their buildings and hide, scoring a tie.

SC2LE also provides "mini games", similar to [38], that are small scenarios focused on subtasks in the game such as controlling units, gathering resources and building units and structures. They are created using the SC2 map editor, meaning it is easy to share scenarios and create new ones. There are seven mini games provided by default, each increasing difficulty, and [40] use the same methods as for the full game mentioned above for these, comparing results with a "novice" player and a grandmaster (top rank on the SC2 leaderboards). No method could match the grandmaster but in the easiest

[1] https://github.com/MartinRooijackers/LetaBot

[2] https://github.com/bluechill/Soar-SC

[3] https://github.com/bwapi/bwapi

four mini games the novice was beaten by the artificial agents. This leaves the three more difficult mini games as a challenge.

Another advantage with SC2LE is that replays can be processed, at speed, through the same interface as agents use. This allows for "bootstrapping" learning in the above mentioned systems by observing replays, especially since Blizzard releases thousands of new replays from ranked games each day. [40] did this and even though the agents were trained on all matchups on different maps of the full game, they were able to increase performance in both the full game and the mini games.

4 BANDITS, RL AND TASK STRUCTURE

Seeing as SC2LE provides an "as human as possible" approach to the agent interface and the possibility of creating custom mini games that are easily shared, there is an opportunity to use this framework as a laboratory environment to test hypotheses on learning and decision making. The fact that SC2 is a game played by millions and not something created specifically for experiments in the cognitive sciences supports a novel and interesting tradeoff between internal and external validity [22]. This would be valuable for informing development of artificial agents both in research and in industry for improved player modelling. If we can replicate previous results from the cognitive sciences it would open the door to more confidently using results from abstract tasks in this more dynamic domain. And if previous results are not replicated, a whole new line of research can be opened instead, investigating what aspects of this new domain makes it different from more strict yet intrinsically less engaging laboratory tasks.

One very commonly used task, with a rich history in the cognitive sciences, is the Iowa Gambling Task [1] and its many variations. In computer science and reinforcement learning especially, analogs of this kind of decision task are called bandit problems [35]. An interesting variation of the task is probabilistic reversal learning tasks [7, 15] where the optimal choice changes according to some probabilistic function and subjects need to discover the pattern in order to maximise their long term reward. [13] show that if participants are aware of this higher order task structure, RL models cannot account for blood flow signal changes in prefrontal cortex as well as a Bayesian hidden Markov model. This is perhaps because RL models update action values one at a time, whereas humans seemingly can have "aha" moments [37] and change their expectations of all options simultaneously [8, 13]. This phenomenon is theorized to be an effect of interactions between the above mentioned System 1 (RL in the basal ganglia dopamine system) and 2 (higher level functions, likely in the human prefrontal cortex) [12, 16, 19, 43].

5 OUR STUDY

The study we propose is thus based on the simplest mini game provided with SC2LE; MoveToBeacon, where the goal is to move a single marine on top of a green beacon to receive a reward. We extend this task with a version of the task used in [13, 45] in the form of an additional blue beacon, creating a two arm bandit task. During eighty trials, the beacon providing the optimal reward will switch at certain times, so participants need to adjust their behaviour in response to the changed contingencies. When the marine steps on a beacon, both beacons reappear at new points on the screen equidistant from the marine in order to keep the two options as equal as possible with regards to saliency and effort.

We aim to recruit at least one hundred[4] participants via online message boards and chat groups. They will be directed to a questionnaire, which gathers consent, basic demographic information and extraversion score based on EPQ-BV [29]. Because players have different levels of experience with the game, we ask how long they have played the game, if they are ranked and if so, what rank. Then follows instructions on how to play the game and upload their replay for our analyses.

Despite the potential confounds of being able to click anywhere on the screen, and having to wait for the marine to run across the map to the next beacon, we believe it is likely we will replicate previous results from similar studies using lab-based tasks [13, 45]. We further believe the use of SC2 increases motivation in subjects compared to the classical approach of abstract tasks.

Participants' observed choices in this task will be analysed by fitting models to the data. For RL, we will focus on Q-learning and Sarsa, and sub-variants, as these have the strongest biological evidence [25] and is also supported by [13]. Additionally, the state inference model from [13] will also be used, and the degree of fit for all models will be compared as per [9, 13]. Our hypothesis is that those models equipped with state inference will fit the data better than the RL ones. We are currently collecting data and expect to have results in the coming months.

6 DISCUSSION & FUTURE WORK

After modelling the human behaviour in our task described above we aim to use these results to inform algorithms and potentially infer parameter value ranges to create an artificial agent that replicates the same task. This agent would likely receive the beacon positions as specific actions, to make more direct comparisons to human behaviour. This is because end-to-end approaches often overlook that object identification is integrated with the RL system. For example, in the MoveToBeacon task a human would immediately identify the marine and the beacon and quickly infer they can make the two interact. Artificial agents, meanwhile, need many trial and error attempts to identify the objects involved and the action sequences leading to high rewards. So in practice, end-to-end approaches are more comparable to newborn humans than adults.

The foregoing point leads us to another interesting question, given that humans who have never played *StarCraft* before still have much knowledge that helps them understand how to play. Thus, we can ask how much "pre-programmed" knowledge does an artificial agent need to be comparable? This question of innate knowledge is an active area of research with some claiming a high degree and sophistication of this knowledge [34] and some arguing it is more akin to an emergent phenomenon [33]. Lately the question has been getting increased attention within AI [21].

These questions are not meant to take anything away from the amazing achievements of deep reinforcement learning, but

[4] In previous pilot research with bandit tasks with switching rewards (using task from [45]), we found that the fitted learning rate parameter was inversely and significantly correlated with extraversion after partialling out the fitted inverse temperature parameter controlling response choices (partial $R^2 = 0.1$; effect size $f^2 = 0.11$) . For a two-tailed test of this partial correlation, at $p = 0.05$, in order to achieve 90% power to detect such a relationship, G*power [10] reveals that 98 participants are required.

rather point out how incredible humans are. This is what makes SC2LE and its "human-like" approach to the agent interface so exciting. It provides a straightforward way to create a plethora of experiments that can have both humans and artificial agents as participants and results from both feeding into each other, with the additional possibility of examining replays of ranked matches for additional data. This is valuable not only for our knowledge of human learning and decision making but also for creating models of player behaviour, new methods for GVGP and thusly also artificial general intelligence.

REFERENCES

[1] A Bechara, A R Damasio, H Damasio, and S W Anderson. 1994. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50, 1-3 (April 1994), 7–15.

[2] Blizzard Entertainment. 1998. Starcraft: Brood War.

[3] Blizzard Entertainment. 2010. Starcraft II.

[4] Michael Buro. 2003. Real-time strategy games: A new AI research challenge. In *IJCAI*. 1534–1535.

[5] Michal Certicky and David Churchill. 2017. The Current State of StarCraft AI Competitions and Bots. In *Workshop on Artificial Intelligence for Strategy Games*.

[6] David Churchill, Mike Preuss, Florian Richoux, Gabriel Synnaeve, Alberto Uriarte, Santiago Ontañnón, and Michal Čertický. 2016. StarCraft Bots and Competitions. In *Encyclopedia of Computer Graphics and Games*, Newton Lee (Ed.). Springer International Publishing, Cham, 1–18.

[7] Roshan Cools, Luke Clark, Adrian M Owen, and Trevor W Robbins. 2002. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J. Neurosci.* 22, 11 (June 2002), 4563–4567.

[8] Vincent D Costa, Valery L Tran, Janita Turchi, and Bruno B Averbeck. 2015. Reversal learning and dopamine: a bayesian perspective. *J. Neurosci.* 35, 6 (Feb. 2015), 2407–2416.

[9] Nathaniel D Daw. 2011. Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII* 23 (2011), 3–38.

[10] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 2 (May 2007), 175–191.

[11] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2017. Counterfactual Multi-Agent Policy Gradients. (May 2017). arXiv:cs.AI/1705.08926

[12] Jan Gläscher, Nathaniel Daw, Peter Dayan, and John P O'Doherty. 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 4 (May 2010), 585–595.

[13] Alan N Hampton, Peter Bossaerts, and John P O'Doherty. 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* 26, 32 (Aug. 2006), 8360–8367.

[14] Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. 2017. Neuroscience-Inspired Artificial Intelligence. *Neuron* 95, 2 (July 2017), 245–258.

[15] A Izquierdo, J L Brigman, A K Radke, P H Rudebeck, and A Holmes. 2017. The neural basis of reversal learning: An updated perspective. *Neuroscience* 345 (March 2017), 12–26.

[16] Daniel Kahneman. 2011. *Thinking, fast and slow.* Macmillan.

[17] John Laird and Michael VanLent. 2001. Human-level AI's killer application: Interactive computer games. *AI magazine* 22, 2 (2001), 15.

[18] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. 2017. Building machines that learn and think like people. *Behav. Brain Sci.* 40 (Jan. 2017), e253.

[19] Sang Wan Lee, Shinsuke Shimojo, and John P O'Doherty. 2014. Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81, 3 (Feb. 2014), 687–699.

[20] John Levine, Clare Bates Congdon, Marc Ebner, Graham Kendall, Simon M Lucas, Risto Miikkulainen, Tom Schaul, and Tommy Thompson. 2013. General Video Game Playing. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany.

[21] Gary Marcus. 2018. Innateness, AlphaZero, and Artificial Intelligence. (Jan. 2018). arXiv:cs.AI/1801.05667

[22] Arthur B Markman. 2018. Combining the Strengths of Naturalistic and Laboratory Decision-Making Research to Create Integrative Theories of Choice. *J. Appl. Res. Mem. Cogn.* 7, 1 (March 2018), 1–10.

[23] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 7540 (Feb. 2015), 529–533.

[24] Matthew Molineaux, David W Aha, and Philip Moore. 2008. Learning Continuous Action Models in a Real-Time Strategy Environment. In *FLAIRS Conference*, Vol. 8. aaai.org, 257–262.

[25] Yael Niv. 2009. Reinforcement learning in the brain. *J. Math. Psychol.* 53, 3 (2009), 139–154.

[26] Santiago Ontanon, Gabriel Synnaeve, Alberto Uriarte, Florian Richoux, David Churchill, and Mike Preuss. 2013. A Survey of Real-Time Strategy Game AI Research and Competition in StarCraft. *IEEE Trans. Comput. Intell. AI Games* 5, 4 (2013), 293–311.

[27] Peng Peng, Ying Wen, Yaodong Yang, Quan Yuan, Zhenkun Tang, Haitao Long, and Jun Wang. 2017. Multiagent Bidirectionally-Coordinated Nets: Emergence of Human-level Coordination in Learning to Play StarCraft Combat Games. (March 2017). arXiv:cs.AI/1703.10069

[28] Glen Robertson and Ian Watson. 2014. A Review of Real-Time Strategy Game AI. *AI Magazine* 35, 4 (2014), 75.

[29] Toru Sato. 2005. The Eysenck Personality Questionnaire Brief Version: factor structure and reliability. *J. Psychol.* 139, 6 (Nov. 2005), 545–552.

[30] Tom Schaul, Julian Togelius, and Jürgen Schmidhuber. 2011. Measuring Intelligence through Games. (Sept. 2011). arXiv:cs.AI/1109.1314

[31] Wolfram Schultz. 2015. Neuronal Reward and Decision Signals: From Theories to Data. *Physiol. Rev.* 95, 3 (July 2015), 853–951.

[32] A Shantia, E Begue, and M Wiering. 2011. Connectionist reinforcement learning for intelligent unit micro management in StarCraft. In *The 2011 International Joint Conference on Neural Networks*. IEEE, 1794–1801.

[33] Sylvain Sirois, Michael Spratling, Michael S C Thomas, Gert Westermann, Denis Mareschal, and Mark H Johnson. 2008. Précis of neuroconstructivism: how the brain constructs cognition. *Behav. Brain Sci.* 31, 3 (June 2008), 321–31; discussion 331–56.

[34] Elizabeth S Spelke and Katherine D Kinzler. 2007. Core knowledge. *Dev. Sci.* 10, 1 (Jan. 2007), 89–96.

[35] Richard S Sutton and Andrew G Barto. 1998. *Reinforcement learning: An introduction.* Vol. 1. MIT press Cambridge.

[36] Gabriel Synnaeve, Nantas Nardelli, Alex Auvolat, Soumith Chintala, Timothée Lacroix, Zeming Lin, Florian Richoux, and Nicolas Usunier. 2016. TorchCraft: a Library for Machine Learning Research on Real-Time Strategy Games. (Nov. 2016). arXiv:cs.LG/1611.00625

[37] Martin Tik, Ronald Sladky, Caroline Di Bernardi Luft, David Willinger, André Hoffmann, Michael J Banissy, Joydeep Bhattacharya, and Christian Windischberger. 2018. Ultra-high-field fMRI insights on insight: Neural correlates of the Aha!-moment. *Hum. Brain Mapp.* (April 2018).

[38] A Uriarte and S Ontanón. 2015. A benchmark for starcraft intelligent agents. *Eleventh Artificial Intelligence and Interactive Digital* (2015).

[39] Nicolas Usunier, Gabriel Synnaeve, Zeming Lin, and Soumith Chintala. 2016. Episodic Exploration for Deep Deterministic Policies: An Application to StarCraft Micromanagement Tasks. (Sept. 2016). arXiv:cs.AI/1609.02993

[40] Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John Agapiou, Julian Schrittwieser, John Quan, Stephen Gaffney, Stig Petersen, Karen Simonyan, Tom Schaul, Hado van Hasselt, David Silver, Timothy Lillicrap, Kevin Calderone, Paul Keet, Anthony Brunasso, David Lawrence, Anders Ekermo, Jacob Repp, and Rodney Tsing. 2017. StarCraft II: A New Challenge for Reinforcement Learning. (Aug. 2017). arXiv:cs.LG/1708.04782

[41] Jane X Wang, Zeb Kurth-Nelson, Dharshan Kumaran, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Demis Hassabis, and Matthew Botvinick. 2018. Prefrontal cortex as a meta-reinforcement learning system. *Nat. Neurosci.* (May 2018).

[42] Stefan Wender and Ian Watson. 2012. Applying reinforcement learning to small scale combat in the real-time strategy game StarCraft:Broodwar. In *2012 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 402–408.

[43] Robert C Wilson, Yuji K Takahashi, G Schoenbaum, and Yael Niv. 2014. Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81, 2 (Jan. 2014), 267–279.

[44] S Wintermute, J Xu, and J E Laird. 2007. SORTS: A human-level approach to real-time strategy AI. *Ann Arbor* (2007).

[45] Darrell A Worthy, W Todd Maddox, and Arthur B Markman. 2007. Regulatory fit effects in a choice task. *Psychon. Bull. Rev.* 14, 6 (Dec. 2007), 1125–1132.