Explore-exploit behaviors predict broad autism social phenotypes in general population

A THESIS
SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA
BY

Evan Knep

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

Advised by Dr. Nicola Grissom

2024

Abstract:

Autism spectrum disorder (ASD) is most often defined by two major behavioral domains, social communication challenges and behavioral rigidity, but executive function deficits have long been considered potential contributors to autism-related impairments across these domains. Autism is a spectrum, with a broad range of autism-related behaviors presenting below the diagnostic threshold, raising the interesting possibility that variability in executive function may contribute to individual differences in autistic traits in the general population. Value based decision making tasks access aspects of executive function, and critically are amenable to computational approaches to dissect the latent variables that most contribute to individual differences in cognition. We capitalize on this approach to uncover the relationship between autistic traits as measured by the Broad Autism Phenotype Questionnaire (BAPQ) and explore-exploit balance in a three-armed restless bandit decision making task in a large (1001 participants) sample. We find that the BAPQ aloof subscale, which primarily describes social behavior related phenotypes, most strongly explains changes in choice behaviors, including sensitivity to outcomes, changes in choice flexibility, and level of exploration as inferred from a Hidden Markov Model. Canonical correlation analysis reveals that the strongest loading for these non-social reward related measures are in fact socially coded items. These findings suggest that different aspects of executive function challenges may be related to social and nonsocial autism-related behaviors in the general population, and that social components of behavior produce measurable differences in nonsocial tasks.

Table of Contents

Introduction

The autism spectrum represents a broad range of phenotypes linked by alterations in several core behavioral domains, namely behavioral rigidity and social communication challenges (*Diagnostic and Statistical Manual of Mental Disorders: DSM-5™, 5th Ed*, 2013). A long standing question in autism research is what might be linking these seemingly discrete behavioral domains, such that alterations in both consistently emerge within autistic individuals (Caronna et al., 2008; Happé, 2001). Recent evidence has implicated executive function as a key contributor to autism-related phenotypes across behavioral domains (Craig et al., 2016; Gilotty et al., 2002; Grissom & Reyes, 2019; Leung et al., 2016; Russell, 1997). Dysfunction in goal-directed behaviors, behavioral flexibility, or habit breaking behaviors all present very real challenges in a complex and unpredictable world, and could underpin both behavioral stereotypy and difficulties integrating social cues into social behavior. Though there is considerable evidence that challenges with executive function are prevalent in ASD diagnosed individuals, it remains unclear how executive functions may be differentially contributing to autistic traits across each core behavioral domain of autism.

One cognitive process in the domain of executive function that is altered in ASD is reward-guided decision making (Grissom & Reyes, 2019). People on the autism spectrum show stereotyped challenges in decision making tasks, including difficulty in learning new reward associations and an increase in perseverative errors (Brosnan et al., 2014; Mussey et al., 2015; Solomon et al., 2011, 2015; Vella et al., 2018; Yechiam et al., 2010). One commonly used task to assess decision making is the Multiarmed Bandit (Berry & Fristedt, n.d.; Chen, Ebitz, et al., 2021; Chen, Knep, et al., 2021; Dayan & Daw, 2008; Ebitz et al., 2018). In the bandit task, participants are asked to repeatedly choose between two or more choices whose value is unknown and must be learned through experience. Applying computational models to bandit tasks allow for the dissociation of cognitive components of decision making as latent cognitive axes (Chen,

Ebitz, et al., 2021; Chen, Knep, et al., 2021; Dayan & Daw, 2008; Ebitz et al., 2018; Gershman & Daw, 2017). Most computational models distinguish two major latent cognitive processes: exploiting a rewarding option when it's found and exploring for information about the environment. Balancing exploration and exploitation, collectively a "strategy", is essential for adapting to a changing environment and sustaining long-term reward. Examining individual strategy allows us to link latent cognitive axes of decision making with autism-related phenotypes across behavioral domains.

In this study, we aimed to examine how executive function, as measured through latent cognitive strategies employed to navigate an uncertain environment, relates to broad autism phenotypes across social and nonsocial behavioral domains. To maximize the range of cognitive strategies and autism-related phenotypes captured, we recruited from a large general online sample (n=1001) through the Prolific platform. We used a variant of a classic bandit task, the three-armed restless bandit, to examine how individuals learn and explore in an uncertain environment. The "restless" twist of the task refers to a reward contingency that is not only stochastic, but also dynamic. The reward probability associated with each choice changes randomly and independently over time, requiring subjects to constantly transition between exploiting previously rewarding options and exploring for new information. As the restless bandit task provides no constraints on when and how individuals explore the changing environment, it encourages a wide range of strategies, which provide insight into the latent cognitive processes that may contribute to autism-related phenotypes. To identify the spectrum of autism-related phenotypes in our sample population, we adopted the Broad Autism Phenotype Questionnaire (BAPQ). The BAPQ is ideally suited for measuring subtle phenotype differences as it was designed to measure broad autism phenotypes in non-diagnosed relatives of ASD diagnosed individuals (Hurley et al., 2007; Piven & Sasson, 2014). The BAPQ robustly captures autism phenotypes across three core behavioral domains: social aloofness, behavioral rigidity, and pragmatic language (a measure of language difficulties) . We examined how parameters of decision making strategy varied across the three BAPQ domains. We find that the aloof subscale, which primarily describes

social behavior related phenotypes, most strongly explains changes in choice behaviors, including sensitivity to outcomes, changes in choice flexibility, and level of exploration. A canonical correlation analysis of individual questions in the BAPQ also revealed that the questions with the strongest loading for these presumed non-social behavioral measures are in fact socially coded items. Together, these findings suggest that broad autism phenotypes are associated with individual differences in behavioral measures of executive function, and that the autistic phenotypes most related to these executive function differences may lie within the domain of social communication challenges.

Results:

1001 participants over the age of 18 were recruited using the online platform Prolific to examine the relationship between autism-related phenotypes and latent cognitive processes underlying value-based decision making. To measure expression of autism related phenotypes, we adopted the Broad Autism Phenotype Questionnaire (BAPQ) (Hurley et al., 2007). Participants were asked to complete the questionnaire prior to the behavioral task. The BAPQ is a commonly used tool to quantify autism related traits across three major phenotypic domains, including aloofness, rigidity, and pragmatism (**Figure 1C**). The total BAPQ score ranges from 36-216, with higher scores associated with more severe autism-related phenotypes. The average score of the participant population was 113.98 with a standard deviation of 23.22. The total score consists of three subscores (aloof, rigid, pragmatic), with each subscale score ranging from 12-72. The average score of BAPQ aloof was 37.8 with a standard deviation of 10.59. The average score of BAPQ rigid was 39.83 with a standard deviation of 8.62. The average score of BAPQ pragmatic was 36.16 with a standard deviation of 7.87.

To examine individual differences in value-based decision making strategies, we employed a restless three-armed bandit task. On each trial participants were given three decks of cards to choose from, each of which was associated with a reward probability that changed randomly and independently over time (**Figure 1A-B**).

Participants accrue one point for every rewarded trial. The goal of the task is to maximize the number of points obtained over the duration of the experiment (300 trials), however the participant can only learn about the payoff potential of each deck by randomly sampling and learning about the outcomes. This dynamic reward contingency naturally encourages participants to be flexible in their decisions as the current best option may become worse in the future. To maximize reward, participants must exploit a favorable option when it's found, but also flexibly explore other options to gather information about potential better alternatives. To verify that participants understood the task and were able to sustain reward, we calculated the probability of obtaining reward compared to the probability of reward if choosing randomly (chance). The results suggest that participants were performing the task significantly better than chance (one-sample t test, $t(1000) = 79.14$; $p<0.0005$) (**Figure 1D**).

BAPQ Aloof Scores Drive Decreased Shift Behaviors

Deficits in cognitive flexibility, a key measure of executive function, are a classic autism phenotype, and the degree to which one can be flexible in their choices is a key component of cognitive strategy. One way to measure cognitive flexibility is to examine shift behavior (Dajani & Uddin, 2015; Stemme et al., 2007). We calculated the probability of shifting to a different option on a given trial and examined whether this shifting behavior correlated with BAPQ total individual subscale scores. We find that the probability of shifting away from a previous choice is correlated with the BAPQ aloof subscale, with high BAPQ aloof scores predicting decreased shift behavior (spearman correlation, $rho=-0.1166$, $p=0.0002$) (**Figure 1E**). BAPQ rigid and pragmatic subscales were not significant predictors of shifting behavior in either direction. These results suggest that the variability in deficits of cognitive flexibility demonstrated across the autism spectrum may be related to social aloofness in particular, highlighting a potential link between social deficits in autism and other cognitive domains.

Aloofness-driven Stay Behaviors are Outcome Dependent

One possible explanation for the decrease in shift behaviors in high aloofness individuals is that they are more likely to repeat as a general strategy, but another possibility is that the decrease in shift is outcome driven, i.e. only shifting when not rewarded. To understand what underlying cognitive strategies may be resulting in decreased shift behaviors in high aloofness individuals, we examined measures of win-stay (repeating the choice given the previous choice was rewarded), lose-shift (shifting to a different choice given the previous choice was not rewarded), and a measure of punishment sensitivity that considers the ratio of shifts as a result of no reward relative to reward while controlling for overall amount of shift behavior. We find a significant positive correlation between BAPQ aloof scores and win-stay (spearman correlation, rho=0.11, p=0.0005), as well as a significant negative correlation between BAPQ aloof scores and lose-shift (spearman correlation, rho=-0.09, p=0.0042) behaviors, which are both consistent with the decreased shift behavior overall with an increase in aloofness (**Figure 1F-G**). To determine whether these results are indicative of sensitivity to reward, no reward, or an artifact of the decreased shift behavior overall, we used punishment sensitivity (**Equation 1,** see methods). We find a significant positive correlation between punishment sensitivity and BAPQ aloof, indicating an increased sensitivity to no reward relative to reward as aloofness increases (**Figure 1H**). This suggests that high aloofness individuals showed decreases in shift behaviors due to higher sensitivity or aversion to no reward, rather than a general strategy of increased choice repetition.

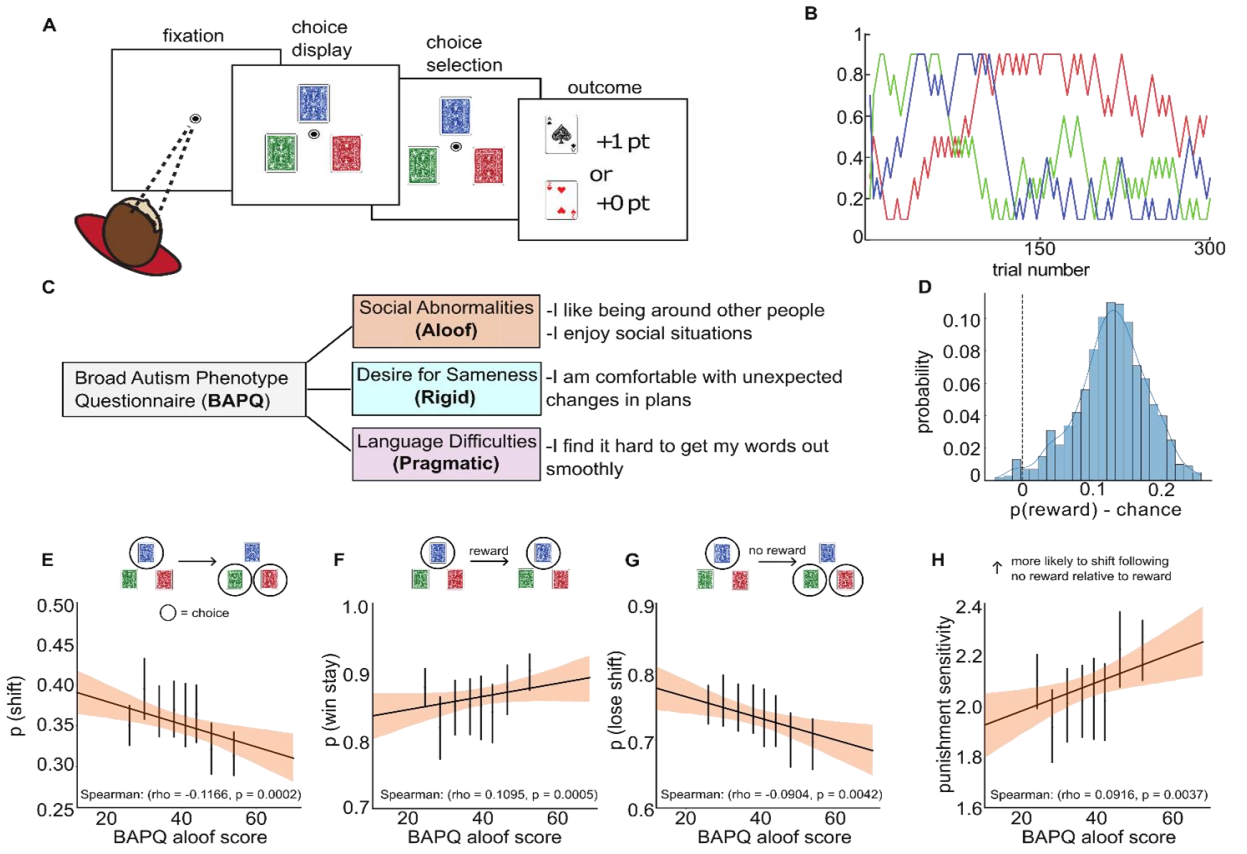

Figure 1. Aloof subscale drives changes in shift behavior

A) Schematic of three-armed bandit task. **B)** Example illustration of the changing reward contingencies across 300 trials. **C)** A self-report version of the Broad Autism Phenotype Questionnaire (BAPQ) was used to assess ASD like behaviors. BAPQ is broken down into three subscales: aloof, rigid, and pragmatic language, with example questions from each category shown. **D)** Probability of obtaining reward relative to chance across participants. Dotted line represents chance. **E)** Probability of shifting on a given trial relative to aloof subscale. **F)** Probability of repeating a choice following reward relative to aloof subscale. **G)** Probability of shifting from previous choice following an unrewarded choice relative to aloof subscale. **H)** Sensitivity to unrewarded trials relative to rewarded trials, normalized to overall shiftiness of the participant, relative to aloof subscale.

BAPQ Aloof is Correlated with Decreased Frequency of Exploratory Behaviors

Changes in outcome-dependent behaviors such as win-stay and lose-shift could be a manifestation of changes in strategy employed to explore the changing environment (Chen, Knep, et al., 2021; Ebitz et al., 2018). Previous studies have shown win-stay and lose-shift behaviors are not constant throughout the session, rather, they are elevated only during periods of exploration (Chen, Knep, et al., 2021; Ebitz et al., 2018). Therefore, we ask whether changes in outcome-dependent behaviors in high aloofness individuals are due to changes in exploratory strategy over time. To infer when exploration happens or how much individuals explore, we adopted a Hidden Markov Model (HMM) that models exploration and exploitation as two latent goal states underlying choices. Previous studies have found that labeling explore-exploit states using an HMM consistently and accurately modeled choice behavior in rodents, non-human primates, and humans (Chen, Knep, et al., 2021; Ebitz et al., 2018; Kaske et al., 2022). To examine how much individuals explore, we calculated the frequency of exploratory choices as labeled by the HMM. This analysis revealed that aloofness was correlated with exploratory behaviors, with high BAPQ aloof scores predicting decreased probability of exploring on a given trial (spearman correlation, rho=-0.135, $p<0.001$) (**Figure 2B**).

Decreased exploratory choices could be a result of decreased frequency of exploratory bouts (less likely to start exploring) and/or shorter exploratory bouts (exploring for fewer trials before settling with a favorable choice). To determine the process that underlie decreased exploratory choices in high aloofness individuals, we examined the transition probabilities between states, i.e how likely an individual was to stop exploiting and start exploring, and how likely an individual was to continue exploring given the previous trial was exploratory (exploration potential). The results revealed that BAPQ aloof correlated with exploration potential but not exploitation potential (continuing to exploit following an exploit trial) (**Figure 2C-D**). Individuals with higher aloofness were less likely to continue exploring after exploring in a previous trial, indicating shorter bouts of exploration,

however they were not increasing the length of their exploitative bouts. This result suggests that the decrease in exploratory behaviors in high aloofness individuals is a product of their committing to a specific option more rapidly, but not sticking to that option for a longer period compared to less aloof individuals.

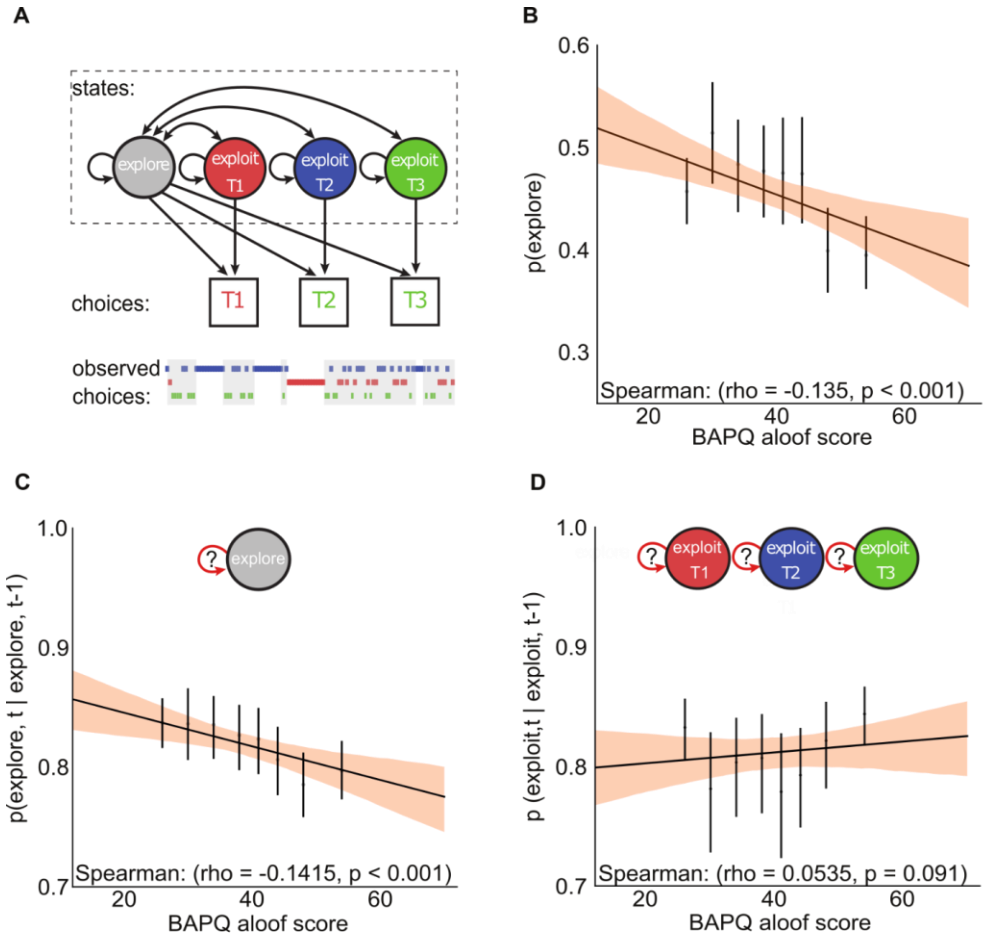

Figure 2. High aloof individuals exhibit decreased exploration

A) Schematic of Hidden Markov Model (HMM) used to label latent goal states underlying choice behaviors. The model incorporates three exploit labels for repetitions of each of three choices, as well as an explore state in which choices were distributed across multiple options. Gray bars on observed choices represent choices labeled as exploratory. **B)** Percentage of trials labeled as exploratory relative to aloof score. **C)**

Probability of exploring following an explore trial relative to aloof score. **D)** Probability of exploiting following an exploit trial relative to aloof score.

Social Behavior Related Questions Show Highest Influence on Exploratory Strategy

Our correlation analyses revealed how differences in exploratory strategy can be predicted by the subscales of BAPQ, specifically BAPQ aloof. Next we ask whether this relationship between exploration and aloofness is driven by one or two specific questions within the aloof subscale or driven by social behavior related questions as a whole. To further investigate which specific aspects within the BAPQ subscales most affected cognitive strategy, we next conducted a canonical correlation analysis to estimate the predictive power of each question. Specifically, we adopted sparse canonical correlation analysis (sCCA), an unsupervised learning algorithm used to maximize the correlations between two datasets with a sparseness parameter control. We chose the sparse solution due to its ability to facilitate the robustness and interpretability of the results (Mihalik et al., 2022; Parkhomenko et al., 2009; Witten et al., 2009)

In this study, the two input datasets consisted of 36 single-item scores from BAPQ (one for each question) and nine behavioral parameters that characterize cognitive strategy, extracted from the restless bandit task (**Figure 3A**).

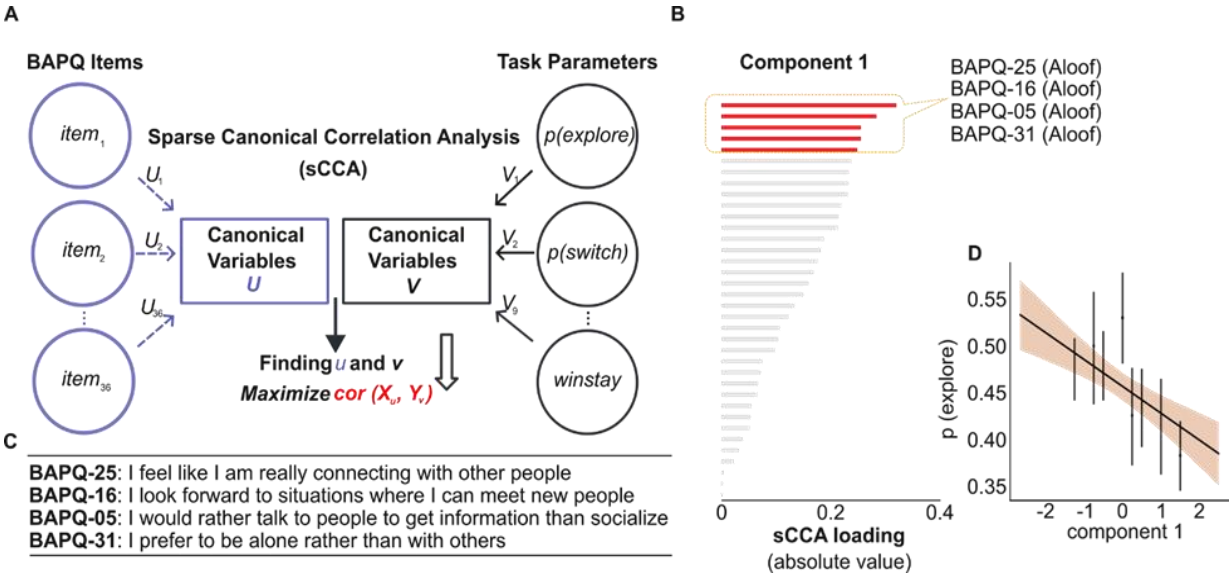

Figure 3. Social specific questions show highest influence on exploratory behaviors in Sparse Canonical Correlation Analysis

A) Schematic of sCCA used to examine effects of specific questions on choice behavior and cognitive strategy **B)** Component one shows highest loading for aloof questions, specifically **C)** Questions with highest loading in component 1, listed in order of highest to lowest **D)** Overall probability of exploration relative to component 1.

For the behavioral parameters, we chose two model-based parameters including (1) probability of exploration, (2) exploration potential inferred from our Hidden Markov Model, and six model-free behavioral parameters including (3) punishment sensitivity, (4) probability of shift away from the previous choice, (5) probability of repeating a rewarded choice (win-stay), (6) probability of shifting away from a non-rewarded choice (lose-shift), (7) relative response time difference (*shift* vs *stay*), (8) averaged response time overall, (9) probability of obtaining a reward controlled chance level.

We conducted sCCA in a predictive framework (Mihalik et al., 2022) to enhance the generalizability of the model. The L1 penalty for the single-item BAPQ dataset and behavioral indices from exploration-exploitation task were tuned by 10-fold cross-

validation, with a discovery (n = 667) and replication sample (n = 334). We obtained the best model with a penalty level of 1 on the single-item BAPQ and 0.9 on the exploration-exploitation behavioral indices.

We selected the first canonical variate based on the degree of covariance explained for further analysis. To test the significance of this canonical variate, we performed a permutation test (permutation 5000 times) as suggested by previous literature (Mihalik et al., 2022; Wang et al., 2020). Null distribution was built by randomly re-assigning subjects' all single items of BAPQ to shuffle the original correlation between two datasets. Results showed that the first canonical variate is significant ($p=0.033$). This analysis suggests that this component is unlikely to have occurred by chance. To explain the psychiatric meaning of this component, we extracted canonical loadings for all single items within BAPQ (**Figure S1**).

Interestingly, the component was dominated by symptoms related to social behaviors, symptoms with the top 10% loading (absolute value) were "I look forward to meeting new people", "I feel like I am really connecting with other people", "I would rather talk to people to get information than socialize", and "I prefer to be alone rather than with others" (**Figure 3B**). On the task side, this component was dominated by the probability of exploratory choice (i.e., p(explore)). Further, we conducted correlation analysis for p(explore) and the total score of these four symptoms, finding a strong correlation between the two ($rho=-0.130$, $p<0.001$) (**Figure 3D**).

Together, our results demonstrate that individual differences in cognitive strategy, a measure of executive function, were most predictive of social-related autism phenotypes in a general population. This suggests that variability in executive function differentially contributes to the behavioral and social domains of autism phenotypes, highlighting the importance of considering the impact of mechanisms underlying social behaviors even in nonsocial tasks.

Discussion

In this study, we examined the relationship between behavioral and social domains of the autism spectrum and how they relate to measures of executive function in a general population. We used computational modeling approaches to extract latent cognitive axes of executive function from a value-based decision making task and found significant differences in latent processes underlying executive function that were primarily driven by the BAPQ aloof subscale. Increased aloofness was strongly correlated with decreased shift behaviors, increased sensitivity to unrewarded trials, and decreased levels of exploration as inferred from a Hidden Markov model (HMM). We also showed that these effects were not driven by a one or two specific questions in the aloof subscale, rather the most predictive questions, as measured by the highest loading in the sparse canonical correlation analysis (sCCA), were all within the aloof subscale. These findings suggest that specific autism-related behavioral domains are differentially associated with measures of executive function, and highlight the influence of mechanisms underlying social behaviors in a nonsocial task.

A major advantage of pairing a self-report questionnaire such as the BAPQ with a decision making task such as the restless bandit is the ability to compare *revealed vs reported* trends within the data. In other words, we are able to ask if what people *report* as their behavior matches with what is *revealed* in a task accessing this behavior. To this point, it's interesting that self-reported levels of behavioral rigidity were not reflected in the behavioral strategy in the bandit task. The rigid and pragmatic subscales were not correlated with common measures of behavioral flexibility such as shiftiness, but instead most correlated with social aloofness. It has long been recognized that self-reports are not necessarily accurate with regards to cognitive phenomena, and that similar labels on self report items and task measures do not mean that they access the same construct (Dang et al., 2020). However, it is possible that participants are better able to accurately self-reflect on the aloof subscale prompts, because they access the motivation to perform social behaviors, rather than mere competence (Dang et al., 2020; Saunders et al., 2018). Responding to "I enjoy chatting with people" (from the

aloof subscale) better reflects motivation and preference than responding to questions such as “I can tell when it is time to change topics in conversation” (from the pragmatic subscale), reflecting competence. Because autistic traits have repeatedly been linked with changes in motivation (Chevallier et al., 2012; Mundy, 1995; Schultz, 2012), this may be an important contributor to why the aloof scale in a general population best assesses nonsocial motivation.

Our finding that the aloof subscale was the most predictive of behavioral differences naturally raises the question of why. One potential explanation for this lies in the shared striatal circuitry of social information processing and reward processing (Báez-Mendoza & Schultz, 2013; Bhanji & Delgado, 2014). It is possible that deficits in social behaviors as seen in individuals on the autism spectrum are a result of dysfunction in the same circuitry responsible for general reward and goal-directed behaviors, and would explain our observed correlations between aloofness, or lack of social motivation, and decision strategy, reflecting general motivation in this task. However, although the subscales in the BAPQ are not discrete, we cannot rule out that these aloof-centric results may suggest that the subscales do not properly align with the phenotypic domains they are trying to measure. Ultimately we cannot yet determine the answer to this question, but these results nonetheless highlight the exciting potential for capturing the influence of mechanisms underlying social behaviors using fundamentally non-social tasks.

The BAPQ was developed to measure broad autism phenotypes (BAP) in non-autistic relatives of autistic individuals, with supposedly high sensitivity for detecting BAP when comparing relatives of autistic individuals to non-autistic controls (Bang et al., 2022; Hurley et al., 2007). Although the BAPQ categorizes behaviors into three domains, recent work, including the current study, point to the possibility that shared mechanisms contribute to dysfunction across each domain. Currently, there is an exciting new trend across animal models and human research to adopt tasks that allow access to latent axes that capture interindividual variability in functioning with computational modeling approaches (Chen, Knep, et al., 2021; Ebitz et al., 2018). One of our recent studies

using a similar approach identified latent strategies that convey resilience to stress and depression in health care workers during the Covid-19 pandemic (Kaske et al., 2022). Fine-grained computational models, like the ones we developed here, can reveal *computational phenotypes* that convey vulnerability and resilience in mental illness, identify shared circuits and mechanisms underlying neuropsychiatric features as intervention targets, and inform clinical practice.

Methods

Participants

1001 online participants were recruited through Prolific, a recruiting tool for online experiments. Participation criteria included 1) being at least 18 years of age and 2) all participants must complete the task through desktop computers. These criteria were set to allow for a diverse general population from which to sample, with the device limitation implemented to mitigate any potential differences that could arise due to variation in modality used to complete the task. Participant ages ranged from 18-75+, with an age distribution as follows. 18-24: 473, 25-34: 339, 35-44:114, 45-54:50, 55-64: 18, 65-74: 6, 75+: 1. All participants provided written informed consent. Participants received $3.10 if they completed the entire study, $3.50 if they had high accuracy during the task, and an additional $2.00 for responding to survey questions.

Self-administered Assessments

Participants were asked to complete the Broad Autism Phenotype Questionnaire (BAPQ) (Hurley et al. 2007), as well as supply demographic information including sex assigned at birth, level of education, and household income.

Exploration-Exploitation Paradigm

Exploratory behaviors were measured using a three-arm restless bandit task (Liu et al., 2013; Speekenbrink & Konstantinidis, 2015). Each trial, participants were presented with three choices, each of which was associated with a reward probability that changed

randomly and independently over time. Rewarded trials resulted in participants earning 1 point. The probabilistic nature of the task necessitates exploratory behaviors to monitor changes in reward rates as the task progresses, while encouraging exploitation of a target while the chance of reward is high. Participants were initially screened with a tutorial section, in which they needed to demonstrate basic task proficiency over 15 practice trials prior to beginning the final task. Each participant's restless "walk", or the volatility of the reward probability across 300 trials, was dictated by predetermined parameters for the likelihood of change in reward probability each trial (hazard rate) and the subsequent magnitude of change in reward probability (step size). In this experiment we used a hazard rate of 0.6667 and a step size of +/- 0.2, constrained within the range of [0.1-0.9].

Data Analysis

General analytical techniques

Data analysis was performed with custom PYTHON scripts. Shapiro-Wilks test was performed on relevant variables, with key variables such as BAPQ subscale scores and percent explore/exploit found to be non-normally distributed. The relationship among and between BAPQ scores, explore-exploit metrics, and other variables was subsequently tested with Spearman's correlation tests. All tests assumed an alpha of 0.05. No participants were excluded from our final data analysis.

Punishment Sensitivity

To determine whether shift behaviors were reward driven, i.e. participants only shift when the previous trial was not rewarded, we used a measure of punishment sensitivity to assess each participant's relative shift probability following rewarded or unrewarded trials. If shift behavior is not reward driven we will see punishment sensitivity close to zero.

$$\text{punishment sensitivity} = (p(\text{shift}|\text{loss}) - p(\text{shift}|\text{win})) / p(\text{shift}) \quad \textbf{[1]}$$

Hidden Markov Model (HMM)

In order to examine how much an individual explored in the task, we fit a Hidden Markov model (HMM) to the choice sequence to infer latent explore/exploit state (Chen, Ebitz, et al., 2021; Chen, Knep, et al., 2021; Ebitz et al., 2018). The HMM modeled exploration and exploitation as two latent goal states underlying choices. Each state is defined by a different emission matrix, i.e. the probability of making each choice under each state. When the model is fitted to the actual choice sequence of a subject, it estimates a transition matrix which dictates the probability of transitioning from one state to another over time. Since the HMM assumes a Markovian process, states are time-dependent. The transition matrix is a mapping of past and future states, which describes the 1-time-step probability of transition between every combination of states.

In our model, there were four possible states (3 exploit states for each choice, and one explore state). During each exploit state, the probability of choosing the exploited choice is 1, and 0 for other choices (fixed emission matrix). For example, exploit-choice1 state emits only choice 1. Exploration is modeled as a uniform distribution over choices because the uniform distribution over choices is the maximum entropy distribution of categorical variables. We want to make as few assumptions as we possibly can about the choices they make during exploration. To accurately estimate the parameters of the model with a limited number of trials, the parameters were tied across exploit states such that each exploit state has the same probability of keep exploiting or begin exploring. Transitions out of exploration into exploit states were also tied. The model also assumed that subjects had to go through an explore state in between exploit states, even for a single trial exploration. The model estimates two unique free parameters - the probability of transitioning from exploration to exploitation and the probability of transitioning from exploitation to exploration.

To find the optimized transition matrix for each subject, we fit the model via expectation maximization using the Baum Welch algorithm (Bilmes & Others, 1998). The algorithm was reseeded 10 times to avoid local maxima and find the global maxima. With the HMM transition matrix optimized from subject choice sequences, we then used the Viterbi algorithm to decode latent states from choices, allowing us to label each choice as either exploratory or exploitative.

Sparse canonical correlation analysis

We applied sparse canonical correlation analysis (sCCA), a well established and popular method to find associations across multiple sets of multivariate data. The main goal of sCCA is to find pairs of linear combinations which would maximize the correlations between two datasets with a sparseness parameter control, and how many dimensions are required for those correlations (38). The pairs of linear combinations represent canonical variables, and the correlation between them are defined as canonical correlations.

In the current study, we have two dataset BAPQ single-items (***B***) and Exploration-Exploitation indices (***E***).

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B_{11}} & \cdots & \boldsymbol{B_{1p}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{B_{n1}} & \cdots & \boldsymbol{B_{np}} \end{pmatrix} \qquad \boldsymbol{E} = \begin{pmatrix} \boldsymbol{E_{11}} & \cdots & \boldsymbol{E_{1q}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{E_{n1}} & \cdots & \boldsymbol{E_{nq}} \end{pmatrix} \qquad \textbf{[2]}$$

Where ***n*** represents the number of participants, ***p*** indicates the number of single items of BAPQ, ***q*** indicates the behavioral indices from exploration and exploitation task.

We aimed to find pairs of linear combinations to maximize the correlation between ***B*** and ***E***.

Mathematically, we can have linear combinations like:

X = Bx Y = Ey

and we have:

$$\text{Var (X)} = x^T \sum_{bb} x$$
$$\text{Var (Y)} = y^T \sum_{ee} y$$
$$\text{Cov (X, Y)} = x^T \sum_{be} y$$

Our aim is:

$$Max_{x,y} x^T \sum_{be} y \ \ subject\ to\ constraints\ \|x\|^2 \leq 1, \|y\|^2 \leq 1$$

$$P_1(x) \leq c_1, P_2(y) \leq c_2$$

Where the $P_1(x)$ and $P_2(y)$ are lasso penalty functions (i.e., L1 regularization), and c_1, c_2 should satisfy:

$$1 \leq c_1 \leq \sqrt{p}\ ,\ 1 \leq c_2 \leq \sqrt{q}$$

The value of c_1 and c_2 are chosen by K-fold cross-validation (CV), where the corresponding penalty values are chosen by grid search in increments of 0.1 between 0.1 and 1.0 to identify the combinations of parameters to maximize the Cov (***B****x,* ***E****y*). As previous studies (Wang et al., 2018; Xia et al., 2018) did, we have 10 randomly resampled datasets as replication datasets, each of them consisted of two-thirds of the dataset with full sample. We conducted sCCA in a predictive framework (25) to enhance the generalizability of the model. The L1 penalty for the single-item BAPQ dataset and behavioral indices from exploration-exploitation task were tuned by 10-fold cross-validation, with a discovery (n = 667) and replication sample (n = 334). We obtained the best model with a penalty level of 1 on the single-item BAPQ and 0.9 on the exploration-exploitation behavioral indices. Analyses were implemented in R package PMA (Witten et al., 2009), available at https://rdrr.io/cran/PMA/man/PMA-package.html

Permutation test

Permutation testing was adopted to assess the significance of canonical variates (Mihalik et al., 2022). We constructed null distribution for all canonical components by shuffling the rows of the BAPQ single items and holding the behavioral indices vector constant. Thus the linkage between BAPQ single items and participants' behavioral indices has been broken. We then conducted sCCA using the same regularization parameters on the re-aligned dataset to get canonical variates. If the canonical variates from preserved sample are significant than the canonical variates from permutated sample, we selected these canonical variates for further analysis. We performed permutation 5000 times and applied False Discovery Rate (FDR) to control for type I error due to multiple comparisons. The significance level is set at 0.05.

References:

Báez-Mendoza, R., & Schultz, W. (2013). The role of the striatum in social behavior. *Frontiers in Neuroscience*, *7*, 233.

Bang, P., Strömberg, M., Meera, S. S., & Igelström, K. (2022). Brief Report: The Broad Autism Phenotype in Swedish Parents of Children With and Without Autism Spectrum Conditions. *Journal of Autism and Developmental Disorders*, *52*(10), 4575–4582.

Berry, D. A., & Fristedt, B. (n.d.). *Bandit problems*. Springer Netherlands.

Bhanji, J. P., & Delgado, M. R. (2014). The social brain and reward: social information processing in the human striatum. *Wiley Interdisciplinary Reviews. Cognitive Science*, *5*(1), 61–73.

Bilmes, J. A., & Others. (1998). A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International Computer Science Institute*, *4*(510), 126.

Brosnan, M., Chapman, E., & Ashwin, C. (2014). Adolescents with autism spectrum disorder show a circumspect reasoning bias rather than "jumping-to-conclusions."

Journal of Autism and Developmental Disorders, *44*(3), 513–520.

Caronna, E. B., Milunsky, J. M., & Tager-Flusberg, H. (2008). Autism spectrum disorders: clinical and research frontiers. *Archives of Disease in Childhood*, *93*(6), 518–523.

Chen, C. S., Ebitz, R. B., Bindas, S. R., Redish, A. D., Hayden, B. Y., & Grissom, N. M. (2021). Divergent Strategies for Learning in Males and Females. *Current Biology: CB*, *31*(1), 39–50.e4.

Chen, C. S., Knep, E., Han, A., Becket Ebitz, R., & Grissom, N. M. (2021). Sex differences in learning from exploration. In *bioRxiv* (p. 2020.12.29.424773). https://doi.org/10.1101/2020.12.29.424773

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239.

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., de Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric Disease and Treatment*, *12*, 1191–1202.

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571–578.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269.

Dayan, P., & Daw, N. D. (2008). Decision theory, reinforcement learning, and the brain. *Cognitive, Affective & Behavioral Neuroscience*, *8*(4), 429–453.

Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed. (2013). *5*, 5.

Ebitz, R. B., Albarran, E., & Moore, T. (2018). Exploration Disrupts Choice-Predictive Signals and Alters Dynamics in Prefrontal Cortex. *Neuron*, *97*(2), 450–461.e9.

Gao, C., Ma, Z., Ren, Z., & Zhou, H. H. (2015). Minimax estimation in sparse canonical correlation analysis. *The Annals of Statistics*, *43*(5), 2168–2197.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*,

68, 101–128.

Gilotty, L., Kenworthy, L., Sirian, L., Black, D. O., & Wagner, A. E. (2002). Adaptive skills and executive function in autism spectrum disorders. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *8*(4), 241–248.

Grissom, N. M., & Reyes, T. M. (2019). Let's call the whole thing off: evaluating gender and sex differences in executive function. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *44*(1), 86–96.

Happé, F. (2001). Social and nonsocial development in autism: Where are the links? In J. A. Burack (Ed.), *The development of autism: Perspectives from theory and research , (pp* (Vol. 374, pp. 237–253). Mahwah, NJ, US.

Hurley, R. S. E., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. *Journal of Autism and Developmental Disorders*, *37*(9), 1679–1690.

Kaske, E. A., Chen, C. S., Meyer, C., Yang, F., Ebitz, B., Grissom, N., Kapoor, A., Darrow, D. P., & Herman, A. B. (2022). Prolonged Physiological Stress Is Associated With a Lower Rate of Exploratory Learning That Is Compounded by Depression. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2022.12.004

Leung, R. C., Vogan, V. M., Powell, T. L., Anagnostou, E., & Taylor, M. J. (2016). The role of executive functions in social impairment in Autism Spectrum Disorder. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *22*(3), 336–344.

Liu, H., Liu, K., & Zhao, Q. (2013). Learning in a Changing World: Restless Multiarmed Bandit With Unknown Dynamics. *IEEE Transactions on Information Theory / Professional Technical Group on Information Theory*, *59*(3), 1902–1916.

Mihalik, A., Chapman, J., Adams, R. A., Winter, N. R., Ferreira, F. S., Shawe-Taylor, J., Mourão-Miranda, J., & Alzheimer's Disease Neuroimaging Initiative. (2022). Canonical Correlation Analysis and Partial Least Squares for Identifying Brain-Behavior Associations: A Tutorial and a Comparative Study. *Biological Psychiatry.*

Cognitive Neuroscience and Neuroimaging, *7*(11), 1055–1067.

Mundy, P. (1995). Joint attention and social-emotional approach behavior in children with autism. *Development and Psychopathology*, *7*(1), 63–82.

Mussey, J. L., Travers, B. G., Klinger, L. G., & Klinger, M. R. (2015). Decision-making skills in ASD: performance on the Iowa Gambling Task. *Autism Research: Official Journal of the International Society for Autism Research*, *8*(1), 105–114.

Parkhomenko, E., Tritchler, D., & Beyene, J. (2009). Sparse canonical correlation analysis with application to genomic data integration. *Statistical Applications in Genetics and Molecular Biology*, *8*, Article 1.

Piven, J., & Sasson, N. J. (2014). On the misapplication of the broad autism phenotype questionnaire in a study of autism [Review of *On the misapplication of the broad autism phenotype questionnaire in a study of autism*]. *Journal of Autism and Developmental Disorders*, *44*(8), 2077–2078.

Russell, J. (Ed.). (1997). *Autism as an executive disorder*. *313*. https://psycnet.apa.org/fulltext/1998-07445-000.pdf

Saunders, B., Milyavskaya, M., Etz, A., Randles, D., & Inzlicht, M. (2018). Reported Self-control is not Meaningfully Associated with Inhibition-related Executive Function: A Bayesian Analysis. *Collabra. Psychology*, *4*(1), 39.

Schultz, R. (2012, May 1). *Social motivation, reward and the roots of autism*. Spectrum | Autism Research News. https://www.spectrumnews.org/opinion/viewpoint/social-motivation-reward-and-the-roots-of-autism/

Solomon, M., Frank, M. J., Ragland, J. D., Smith, A. C., Niendam, T. A., Lesh, T. A., Grayson, D. S., Beck, J. S., Matter, J. C., & Carter, C. S. (2015). Feedback-driven trial-by-trial learning in autism spectrum disorders. *The American Journal of Psychiatry*, *172*(2), 173–181.

Solomon, M., Smith, A. C., Frank, M. J., Ly, S., & Carter, C. S. (2011). Probabilistic reinforcement learning in adults with autism spectrum disorders. *Autism Research: Official Journal of the International Society for Autism Research*, *4*(2), 109–120.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*(2), 351–367.

Stemme, A., Deco, G., & Busch, A. (2007). The neuronal dynamics underlying cognitive flexibility in set shifting tasks. *Journal of Computational Neuroscience*, *23*(3), 313–331.

Vella, L., Ring, H. A., Aitken, M. R., Watson, P. C., Presland, A., & Clare, I. C. (2018). Understanding self-reported difficulties in decision-making by people with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, *22*(5), 549–559.

Wang, H.-T., Smallwood, J., Mourao-Miranda, J., Xia, C. H., Satterthwaite, T. D., Bassett, D. S., & Bzdok, D. (2020). Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. *NeuroImage*, *216*, 116745.

Witten, D. M., Tibshirani, R., & Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* , *10*(3), 515–534.

Yechiam, E., Arshavsky, O., Shamay-Tsoory, S. G., Yaniv, S., & Aharon, J. (2010). Adapted to explore: reinforcement learning in Autistic Spectrum Conditions. *Brain and Cognition*, *72*(2), 317–324.