Social Decision Preferences for Close Others are Embedded in Neural and Linguistic Representations

João F. Guassi Moreira*, L. Concepción Esparza, Jennifer A. Silvers & Carolyn Parkinson

Department of Psychology, University of California, Los Angeles

*Corresponding Author
4572 Pritzker Tower
502 Portola Plaza
Los Angeles CA, 90095
jguassimoreira@ucla.edu

Authors Contributions

JFGM and CP developed research questions and designed the study concept. Data were collected by JFGM (both studies) and LCE (Study 1) under the supervision of CP (Study 1) and JAS (Study 2). JFGM cleaned the data, verified by LCE. JFGM analyzed the data under supervision of CP with feedback from JAS. JFGM drafted the manuscript with substantial feedback and critical edits from CP. LCE and JAS also provided feedback and edits on a later draft. All authors approved the final manuscript for submission.

Abstract

Humans frequently make decisions that impact close others. Prior research has shown that people have stable preferences regarding such decisions and maintain rich, nuanced mental representations of their close social partners. Yet, if and how such mental representations shape social decisions preferences remains to be seen. Using a combination of functional magnetic resonance imaging (fMRI) and natural language processing (NLP), this study investigated how neural and linguistic representations of close others influence social decision-making. Participants nominated a parent and a friend, rated their characteristics, and made hypothetical social decisions while undergoing fMRI. Neural representations of parents and friends, relative to the self, predicted social decision preferences. Specifically, greater neural similarity between self and parent in the temporoparietal junction (TPJ) and nucleus accumbens (NAcc) was associated with a preference for parents, while greater self-friend similarity in the medial prefrontal cortex (mPFC) predicted friend-preference. Additionally, linguistic analysis of written descriptions of close others revealed that social decision preferences could be reliably predicted from semantic features of the text. The convergence of neural and linguistic data further strengthened the prediction of social decision preferences. These findings elucidate the neural and linguistic underpinnings of social decision-making, emphasizing the critical role of mental representations in guiding choices involving familiar others.

Introduction

Humans are frequently faced with decisions that could impact multiple other people. For instance, one may need to choose between having dinner with a close friend or attending happy hour with a new co-worker, or whether to buy a plane ticket home for the holidays to see family or use that money to take a trip abroad with a romantic partner. Decision points such as these are pervasive and often highly consequential for one's relationships with others. Prior work has shown that individuals have specific, stable preferences that guide their choices regarding familiar others[1]. Given that no two people we know are alike, one might expect representations of familiar others (i.e., one's mental model of each person) to play a key role in shaping social decision behavior. However, surprisingly little work has examined this. In the current study, we used functional magnetic resonance imaging (fMRI) and natural language processing (NLP) to examine the link between mental representations of familiar others and social decision behavior towards those individuals.

By and large, most social decision-making studies have not examined how mental representations of familiar others shape behavior. Instead, studies typically investigate how cognitive or affective heuristics are applied over characteristics of choice alternatives to guide decision behavior. These are heuristics that help process a variety of decision-level features that convey information about risk, mood, ambiguity, inequity, harm, others' perspectives, and so on[2–9]. Because so much attention has been devoted to such heuristics and their role in parsing decision-level features of choice alternatives, studies tend to disproportionately focus on contexts in which one's choices impact strangers, anonymous others, and social agents with experimentally manipulated roles[10], despite the fact that everyday social decisions frequently implicate familiar others. Thus, the relationship between our mental *representations* of other people—especially known, familiar others—and the decisions we make regarding them remains largely unexplored.

Available evidence suggests that mental representations of other people should play a role in the decisions we make regarding those people. Several studies have delineated the neural architecture for thinking about other people[11–14], and recent work has further extended this by investigating how distributed neural response patterns encode the identities and attributes of specific others[12,13,15]. This encoding constitutes the basis of mental representations of familiar others – i.e., models that allow us to simulate, forecast, and extrapolate the opinions, thoughts, feelings, and actions of others. Mental representations of others are thought to be evoked when predicting an individual's thoughts or behaviors after having learned about their personality traits[12,16,17], demonstrations of reward or punishment[3,18], and adherence to norms and conventions[19,20], all suggesting that mental representations of other people ought to at least partly drive social decisions. Despite this work, however, it remains unclear how such mental representations influence social choice preferences involving familiar others.

Here we probed if and how mental representations of familiar, close others predict our behavior towards those people. We deliberately chose to investigate mental representations of familiar *close* others for several reasons. First, close others are most commonly implicated in social decision scenarios. Results obtained by investigating close others will be widely applicable to everyday human life – more so than a confederate, fictive individual, or celebrity. Second, relationships with close others can provide much richer representational content to map onto behavior. Third, examining choice preferences among individuals *within* the category

of close others helps map behavioral tendencies in a finer-grained manner (as opposed to coarsely examining choice preferences between close versus non-close others) that is likely to capture the kinds of distinctions that pervade everyday life, where people must often consider the consequences of their choices for multiple familiar others (rather than deciding, for example, whether to favor a friend or a stranger). We specifically studied choice preferences between a parent and close friend because group-level choice preferences between parents and friends have been well documented[1,21,22], and because these two particular relationship categories are motivationally salient across individuals and for much of the lifespan[23–27].

We examined representations using multiple modalities of measurement by using neuroimaging (specifically, fMRI) and applying NLP techniques to written text data. With fMRI, we based our analyses on key theories in social neuroscience that stipulate that one's representation of the self serves as the template for all other mental models, and that self-relevance (i.e., high self-other representational overlap) is a key driver of behavior[16,28]. Thus, we examined the degree to which self-other representational overlap differed between parent and friends in brain regions related to social cognition and value-based processes, and then tested whether this self-other representational overlap tracked with social choice preferences. In particular, the bilateral temporoparietal junction (TPJ) and dorsomedial prefrontal cortex (dmPFC) were probed as regions sensitive to social cognitive processing[29,30] and the medial prefrontal cortex (mPFC) and bilateral nucleus accumbens (NAcc) were specified as value-based processing regions[31].

Using the linguistic data, we employed machine learning to test whether we could engineer a linguistic signature of social decision preferences from written text about one's parent and friend[32], essentially decoding choice preferences from one's representation of others that is available in written text data. Finally, we combined the neuroimaging and linguistic data to test whether convergence of representational content in the brain and language resulted in stronger choice preferences. We did so for two related reasons. First, information from one modality can help disambiguate findings from another. Second, generally, information captured with different methods may capture different aspects of one's mental representations of other people, and using multiple modalities to measure mental representations of others can confer greater sensitivity to links between such representations and behavior. A schematic overview of the entire study's workflow is depicted in Figure 1. We first used neuroimaging data to test for differences in how the brain constructs representations for one's parent and friend using the self as a reference (i.e., are friends or parents represented more similarly to the self?). We anchored representations in relation to the self because of prior work in psychology and neuroscience showing that humans often use mental representations of the self as the basis for representations of others[16,33–35], and that perceived similarity to oneself is often liked to preferences regarding others[9,36,37]. We then examined whether individual differences in the degree of representational similarity between oneself and particular close others was related to social decision preferences involving those individuals. We then used linguistic data to examine whether social decision preferences regarding close others could be predicted from written descriptions of them. To do so, we validated a linguistic signature of choice preferences across several datasets with different types of social decision tasks and with additional characterizations of relationships (e.g., measures of relationship quality). Afterwards, we examined whether information from both modalities (neural, linguistic) converged by correlating

neural representations with linguistic signature scores. Finally, we finished by testing whether high convergence between the two types of representations was related to strong social decision preferences for a given other (i.e., if neural and linguistic representations show a strong parent 'bias', is one more likely to evince a parent-over-friend social choice preference?). In all, this study aims to reveal how mental representations of specific, close others influence social decision behavior towards those individuals.

Results

Information about social decision preferences is embedded in neural representations of close others.

Overview of Approach: Probing Neural Representations of Close Others. We conducted an fMRI study (Study 1) to probe whether and how social decision preferences are embedded in neural representations of known others. Participants in Study 1 (*N* = 63) nominated two close others (one parent, one friend) and completed a social preferences and judgments task about themselves and these others while undergoing fMRI. Across four runs, participants made ratings about how well a set of 75 behaviors, preferences, and attitudes described themselves, their parent, and their friend. On each trial, participants used a button to indicate how well a given item (e.g., "feels comfortable talking about personal issues") described either themselves, their parent, or their friend (1 = "Strongly disagree", 4 = "Strongly agree"). Participants then completed two runs of a social decision-making task where they were required to allocate hypothetical financial (money) or social (time spent) resources between the two close others.

Behavioral ratings and fMRI responses from the preferences and judgments task were used to construct neural representational dissimilarity matrices (RDMs; Figure 1). To examine differences in self-other overlap between parents and friends, we assessed the similarity of participants' representations of themselves to representations of each close other (parent, friend). Differences in self-other overlap between close others could reflect differences in the extent to which representations of these close others draw upon self-relevant information in their construction. Technical details and robustness checks can be accessed in the Methods and Supplement.

Different Categories of Close Others are Differentially Represented Across Brain Regions. We compared the degree to which individuals construct mental representations of parents and friends in terms of self-relevance (i.e., representational overlap between self-parent compared to self-friend) in six *a priori* brain regions of interest (Table 1; Figure 2) that are involved in social processing and social decision-making (ROIs visualized in Supplementary Figure 1). We found modest evidence to suggest greater self-other overlap for parents than for friends in bilaterally in the TPJ and NAcc. In contrast, there appeared to be greater self-other overlap for friends than for parents in mPFC.

We ran two exploratory searchlight analyses to determine if brain regions other than the *a priori* ROIs also distinguish between familiar others. The first of these analyses tested for distinct representations of parents vs. friends (i.e., first-order isomorphic similarity) and the second tested for differences in representational overlap between self and parent, and self and friend (second-order isomorphic similarity). Results of the first searchlight suggested that the

precuneus contains distinct representations of close others (Supplementary Figure 2). Exploratory examination of uncorrected results of this analysis revealed more extensive results in the precuneus and several smaller clusters in medial prefrontal cortex (mPFC; see Supplementary Figure 3). The second searchlight yielded no significant results.

A traditional univariate analysis comparing the magnitude of brain activity evoked when answering items for each target ([self – parent], [self – friend], [parent – friend]) is also included in the supplement (Supplementary Figure 4). This analysis revealed robust activity along the cortical midline and TPJ for both the [self – parent] and [self – friend] contrasts. For thoroughness, we analyzed behavior-based representations using the task's Likert scale ratings and found greater self-other overlap for friends than for parents (posterior Cohen's d = 0.50, [0.28, 0.71] 89% highest density credible intervals).

Social Decision Preferences are Encoded in Neural Representations of Close Others. After examining differences in parent and friend representations, we next tested whether social decision preferences involving parents and friends were embedded in their neural representations. We estimated trial-level choice preferences between parents and friends using hierarchical Bayesian modeling. The model term that captures choice preferences was allowed to vary randomly across participants, allowing us to model the ensuing variability in preference with participant-level data about neural representations. Thus, we entered representational similarities between self and other (parent or friend) for each of the six ROIs into four separate models (bilateral ROIs were included in a single model).

For monetary outcomes, we found that greater self-parent similarity in neural representations in the dorsomedial prefrontal cortex (dmPFC) was robustly associated with a greater tendency to favor parents over friends (Table 2). Greater self-friend similarity in the right NAcc was robustly associated with a greater likelihood of favoring friends (Table 3). Curiously, greater self-friend *dis*similarity in the left NAcc was modestly associated with a greater likelihood of favoring friends. For social outcomes, we observed modest evidence that greater self-friend *dis*similarity in mPFC was associated with a greater propensity to favor friends over parents on the task (Table 2). Greater self-parent similarity in the right NAcc was also modestly associated with an increased propensity to favor parents over friends (Table 3).

Social decision preferences are encoded in linguistic representations of close others.

Overview of Approach: Probing Linguistic Representations of Close Others. We examined whether social decision preferences are also embedded in linguistic representations of parents and friends in Study 2. Study 2 involved analyzing existing written text data across three datasets (total N = 2,364 across several subsamples, see Methods), each initially collected as part of a series behavioral experiments examining social decision preferences between parents and friends[1,21]. Although the exact prompt varied slightly across the three datasets, participants in this study were asked to nominate a parent and a close friend, write down information about them (e.g., memories, adjectives or short phrases describing them), completed at least one social decision-making task that pit outcomes for parents and friends against each other, and completed a measure of relationship quality with each close other. To test whether linguistic representations of others include information about choice preferences, we attempted to construct a predictive and generalizable linguistic signature of social decisions

from the written text data. Our rationale was that if we were able to train a model to predict choice preferences from semantic features of the written text data, use the weights from the model to compute a linguistic preference score, and then predict choice preferences in novel samples and contexts, then that would provide compelling evidence that social decision preferences are encoded in the linguistic representations that surface in written text data.

We first began by training a model to predict social decisions from semantic features of written text data in a subset of Study 2 data. This entailed processing and feature engineering of the written text data and then using leave-one-out cross-validation to choose the best feature engineering method, type of model, and model hyperparameters. Technical details and robustness checks are available in the Methods and Supplement. Model weights were used to compute scalar linguistic preference scores, with a relatively greater value indicating relatively stronger parent choice preference during decision-making and a relatively lower value indicating relatively stronger friend choice preference.

Social Decision Preferences are Encoded in Linguistic Representations of Close Others Model weights were next applied to novel data to predict social decision preferences on a dictator task in which participants allocated a finite sum of hypothetical money between their parent and friend (Figure 3). Linguistic preference scores robustly predicted preferences on this task (r = 0.26, 89% HDI = [0.21, 0.31]). Using hierarchical modeling on data from a multi-trial task, we then observed modest evidence that linguistic preference scores predicted social decision preferences between parents and friends in a multi-trial risky decision task (b = 0.02 [-0.06, 0.11]). Preference scores derived from our linguistic signature were also robustly associated with choice preferences between parents and friends in a multi-trial delay discounting task, when both monetary (b = 0.94 [0.72, 1.20]) and social (b = 1.14 [0.91, 1.39]) outcomes were at stake. Finally, we were able to show that linguistic preference scores generalize to robustly predict *attitudes* about one's relationship with their parents (r_{parent} = 0.23 [0.19, 0.27]). Evidence for null effects was observed for the relationship between preference scores and friend relationship quality, however (r_{friend} = 0.00 [-0.04, 0.04]). These findings are visualized in Supplementary Figure 5.

Linguistic and Neural Representations are Correlated, and Jointly Predict Social Decision Preferences. The preceding findings raise the question of whether having both a strong linguistic preference score and extensive overlap between neural representations of oneself and a given other is associated with a stronger social decision preference for that individual. In other words, if someone's representations of others were 'biased' towards a particular other (e.g., one's friend rather than one's parents) in *both* neural and linguistic data, would we also observe even stronger social decision preferences? Such a pattern of results would have several implications for what kinds of information are stored in neural and linguistic representations, and how they are linked with each other and behavior.

To answer this, we first examined the degree of correspondence between linguistic preference scores and neural similarity (Figure 4, Table 4). We found that preference scores implied by the linguistic signature were correlated with similarity between self and parent neural representations in the anticipated direction for three of the six brain regions (mPFC, R NAcc, L TPJ). Neural similarity between self and friend was correlated with linguistic preferences in the anticipated direction for five of the six brain regions (dmPFC, mPFC, L NAcc, L TPJ and R TPJ).

We then examined whether linguistic preference scores moderated the link between neural representational similarity and choice preferences on the discounting task (Figure 5, Supplementary Tables 4-5). The logic behind this analysis is that linguistic representations and neural representations obtained with the methods used here may capture partially distinct aspects of one's mental representations of others. Thus, combining these different modalities could afford greater sensitivity for detecting associations between one's mental representations of others and behavior. In the monetary outcome domain, a relationship in the non-anticipated direction for the L NAcc was observed with linguistic preference scores and neural dissimilarity between self and friend representations. No other effects met our evidentiary criteria or appeared close to meeting them in the monetary outcome domain. In the social outcome domain, we found that linguistic preference scores interacted with self and parent neural dissimilarity in the mPFC and R NAcc to promote stronger social decision preferences for parents if neural similarity was high and linguistic preference scores were high. A similar finding with self and friend neural dissimilarity in the left TPJ was indicated by examining the posterior distribution but did not quite meet our evidentiary criteria.

While these results indicate some degree of convergence between neural and linguistic representations and subsequently suggest neural representations driving behavior are comprised of semantic information, it remained an open question about what kinds of semantic content are indexed in these representations. To unpack this, we conducted a *post-hoc* descriptive analysis to uncover the semantic content driving linguistic preference scores. First, we passed a subset of Study 2 text data (n = 1,641) through the LIWC-22 software to quantify the prevalence of thematic content along 46 dimensions (see Methods). We then correlated the prevalence of each thematic category with linguistic preference scores. This was done separately for parent and friend text. The ensuing correlations were thresholded at r = |.05|, the approximate threshold for statistical significance given the sample size. Results indicate that affective themes (e.g., risk, want, fatigue) in parent text are the primary drivers of linguistic preference scores; cognitive and state-based themes (e.g., causation, mental health) were the primary drives of linguistic preference scores for friend text (Table 5). Additionally, themes in parent text tended to covary more with linguistic preference scores than those in friend text insofar that more parent text themes evinced correlations that met our threshold.

Next, we took the themes identified in Table 5 and extracted them from Study 1 participants' text data, and then correlated their prevalence in that data with self-other neural similarity from ROIs that showed the strongest relationships with linguistic preference scores (Parent: R NAcc, mPFC, dmPFC, L TPJ; Friend: L NAcc, mPFC, dmPFC, L TPJ; see Figures 4-5). These correlations are plotted in Figures 6 (parent results) and 7 (friend results). Affective and state-based themes were most consistently associated with self-parent representational overlap across ROIs, whereas cognitive themes were most consistently associated with self-friend representational overlap across ROIs. Because of potential negations in the text data (e.g., 'I do not feel happy') we encourage readers to primarily interpret the magnitude, rather than the sign, of these correlations.

Discussion

In the current study, we found evidence that neural and linguistic representations of close others contain information about one's social decision preferences involving these others. We first observed that neural representations of parents and friends are differentially encoded in the brain with respect to the self, such that representations of one's parents are more similar to representations of oneself in bilateral NAcc and TPJ, whereas representations of one's friends are more similar to representations of oneself in dmPFC. Individual differences in the encoding 'bias' of these brain regions were consequential for social decision preferences involving close others, as self-other representational similarities in the NAcc, TPJ, and dmPFC tracked with between-individual social decision preferences. Linguistically, written memories and descriptors of parents and friends were used to train a linguistic signature of social decision preferences. When applied to new data, the signature was able to predict choice preferences between parents and friends on novel decision tasks and was even related to other relationship characteristics (e.g., relationship quality). Critically, the scores from the linguistic signature and neural data were related, showing convergence of representations measured with linguistic and neural data. Finally, we found that individual differences in the degree of linguistic-neural convergence were also predictive of social decision preferences. These findings collectively enrich our understanding of how the mental models that humans create for *specific* others influence choices regarding those others.

Existing studies on social decision-making predominantly focus on understanding decision behavior in terms of the cognitive heuristics that individuals apply over situation-level features or how individuals weigh the mental states of others[2–9]. Our findings suggest that an additional driver of social decision preferences stems from how representations of oneself and others are aligned in the brain. We found that self-other similarity in mPFC and dmPFC was associated with choices preferences between a parent and friend. Specifically, greater neural representational overlap between self and parent was associated with a greater likelihood of favoring one's parent-over-friend, whereas *less* self and friend representational overlap (i.e., greater dissimilarity) was associated with stronger friend-over-parent preferences. The NAcc was also predictive of choice preferences, but the direction of the effect varied based on several factors. These varied patterns, paired with existing psychology and neuroscience literature showing that representations of others are linked to a wide range of cognitive, affective, and semantic processes[35,38–40], suggest these brain regions could be storing and accessing different types of representational information and/or associated mental processes. Indeed, descriptive patterns in the written text data appear to provide some evidence for this possibility. That 'encoding biases' towards one's parent or friend in these regions are related to choice preferences implies the possibility of affordances in these representations[41–43], such that the way in which information (here, representations of close others) is encoded might facilitate particular behavioral tendencies during decision-making.

Relatedly, our linguistic findings—the prediction of choice preferences from text data and model weights generalizing to other facets of relationships—point towards a similar concept of 'psychological affordances'. The most readily available mental information when prompted to describe a close other or write down a memory about them (the prompts used to elicit the text data that was used to train the linguistic signature) may also be the most readily available or salient information when making choices that impact said close other. That our decision tasks did not allow for prolonged deliberation between trials means there was a degree of spontaneity

required when responding. Having one's representations of other people contain behavioral affordances is a potentially efficient facet of the psychological architecture that links mental models with behaviors.

The joint success of mental representations measured with linguistic and neural data in predicting decision preferences between close others is particularly interesting with respect to the ability of linguistic information to disambiguate brain-behavior associations. Neuroimaging data can be ambiguous, as the exact operations that the brain is performing cannot always be inferred. However, the close correspondence between the linguistic and neural data can refine hypotheses about the types of information the brain is using. Here, it suggests that representations of close others in implicated brain regions do not entirely house the same information that is contained in the language people use when asked to describe those individuals. Linguistically, neural representational overlap with both parent and friend relationships was associated with motivationally relevant themes. Parent neural representations were more uniquely linked with state-based themes, whereas friend neural representations evinced stronger relationships with social cognitive and mental health related themes. While the results of these analyses are meant to be descriptive and to inform hypotheses for future research, they broadly suggest that representations of specific familiar others are comprised of distinct psychological content.

Future studies could improve and extend on ours in several ways. First, researchers may consider testing other types of known, and not just close, relationships. Understanding how mental representations influence behavior is likely best accomplished by examining such associations across the entire spectrum of psychological and social distance between an individual and the people in their social environment. It would also be helpful to conduct similar studies across different populations to assess the generalizability of these findings. Second, longitudinal work is critical for quantifying the temporal dynamics of how representations are formed and updated with experience, as this is critical for making causal claims. Such avenues would benefit from complementary use of intensive longitudinal methods such as ecological momentary assessment, or incorporating developmental populations that contain individuals who are frequently introduced to completely novel social networks (e.g., incoming college students, middle or high schoolers, etc.) and thus may serve as ideal populations for pursuing this line of work. On a related note, given that mental representations of others are not static entities, another benefit of longitudinal work would be to examine how mental representations of others fluctuate over time and how such fluctuations correspond to social decision-making behavior. Third and last, we believe that future work should incorporate more extensive written text data about close others so as to bolster sensitivity by more extensively sampling one's mental representations of those individuals. Ideally, future studies utilizing intensive longitudinal methods would incorporate written diaries or essay-like prompts at baseline to gather even more text data about close others, or such studies could ethically obtain written records (e.g., text messages) between participants and close others.

Finally, our study speaks to the importance of studying known close others. This study is another entry in a growing body of work that shows individuals have highly granular behavioral choice tendencies involving other individuals[44–47]. Understanding human behavior requires examining the most relevant and meaningful phenomena in our daily experiences. Knowing *how* individuals represent specific others and in turn use these representational distinctions to guide

behavior is critical for understanding much of human behavior as it unfolds outside the laboratory.

Methods

Study 1

Participants. A total of 63 (Mean age = 18.81 years, SD = 10.5 years, 73% female) individuals participated in the current study. Participants were recruited from the University of California, Los Angeles (UCLA) student body by tapping existing internal recruitment pools, classroom advertisements, and an email from the registrar's office. Our *a priori* sample size goal was to recruit and run as many participants as possible during the winter and spring academic quarters (January – June, 2023). In order to be eligible for the study, participants must have been between 18-22 years at the time of enrollment and exhibit no MRI contraindications (e.g., claustrophobia, non-MR safe metallic implants, etc.). All participants were fluent in English. Four participants did not have imaging data available (3 due to software issues preventing data collection/reconstruction[1], and 1 due to participant failure to fully disclose permanent jewelry during safety screening), rendering a sample size of N = 59 for imaging analyses. Demographics for the sample are described in the Supplement. All participants provided consent in accordance with the policies of the UCLA Institutional Review Board. Participants were compensated $40 (USD) upon completion of the study. All data and materials are publicly available on the Open Science Framework (OSF; osf.io/fzds6).

Protocol. Participants completed the study in a one-hour session at UCLA's Staglin One Mind Center for Cognitive Neuroscience. Participants were greeted by an experimenter upon arrival, consented, nominated a parent and close friend (as described in Supplement), received training for the two fMRI tasks, underwent task-based fMRI scanning, and then completed a post-scan survey. Major elements of this procedure are described in greater detail below.

fMRI Acquisition Parameters. fMRI data were collected using a research-dedicated 3T Siemens Prisma scanner and a 32-channel head coil. Functional scan sequences were designed to be consistent with the sequences used by the Human Connectome Project (humanconnectome.org), which uses the University of Minnesota's Center for Magnetic Resonance Research (CMRR) multi-band accelerated EPI pulse sequences (github.com/CMRR-C2P/MB). A high resolution T1 magnetization-prepared rapid-acquisition gradient echo (MPRAGE) structural image was acquired for registration purposes (TR = 1900 ms, TE = 2.48 ms, Flip Angle = 9°, FOV = 256 mm^2, 1 mm^3 isotropic voxels, 208 slices, A >> P phase encoding). Functional runs were comprised of T2*-weighted multiband echoplanar images (TR = 1240 ms, TE = 36 ms, Flip Angle = 78°, FOV = 208 mm^2, 2 mm^3 isotropic voxels, 60 slices, A >> P phase encoding, multiband acceleration factor = 4). Single-band reference images were acquired for each functional run.

[1] Software issues also prevented collection of behavioral data for the social preferences task for one of these four participants.

Behaviors and Preferences Judgment Task. Mental representations of the participant (self), close friend, and parent were elicited during fMRI scanning with a computerized behaviors and preferences judgment task[28,48]. The task entails rating how well a series of preferences describe a given target. On each trial, a label for a given target was displayed at top of the screen ('SELF', 'PARENT', or 'FRIEND'), a preference was displayed in the center of the screen, and a 4-point Likert scale was displayed at the bottom (1 = "Strongly disagree", 4 = "Strongly agree"). Examples of preferences include "really enjoys living in big cities", "thinks modern art is uninteresting", and "feels comfortable talking about personal issues". Participants used a button box to indicate their response. Each trial was presented for 4500 ms, and was followed by a jittered fixation screen consisting of a black '+' in the center of the screen (mean = ~2000 ms, possible range = [200–10165 ms], distributed exponentially). Trials were 'chunked' by target such that participants would rate at least seven consecutive trials about a given target before moving on to another target to help reduce the cognitive load of rapidly switching between targets. To help ensure participants did not miss a transition between chunks of targets, the fixation cue at the center of the screen was displayed in red prior to the start of a new chunk. Participants completed 225 trials total (75/target), spread across four runs. The number of trials per target varied across runs. Each run typically lasted between 345–375 s, given the random jitter. The number of trials for this task was informed by an empirical analysis of a similar open dataset (see Supplement). Preferences for this task were drawn from an open, online pool used in prior studies (https://jasonmitchell.fas.harvard.edu/links/; accessible on our OSF repository at osf.io/sfc6b). Using methods consistent with prior studies (Tamir et al., 2016), we narrowed the 800+ preference pool down to 80 items that were highly and uniquely indicative of Big Five personality traits[49,50] by applying a genetic algorithm to ratings of the items acquired from an independent sample (10 poles with 8 items[2]). Further stimulus optimization details are included in the Supplement. We chose to optimize the items in relation to the Big Five because personality traits are a key component of individual identities. The task was programmed in jsPsych[51].

Social Decision-Making Task. We used a pair of computerized delay discounting tasks to assess social decision preferences between parents and close friends. Our rationale for choosing a discounting task in particular is presented in the Supplement. Participants were asked to make decisions across two separate runs of a delay discounting task. On each trial, participants were presented with two hypothetical scenarios that pitted outcomes for their nominated parent and friend against each other. One scenario involved a relatively immediate, smaller reward and the other involved a relatively delayed, larger reward. The delays could take the value of zero ('Today'), 2 weeks, 4 weeks, and 6 weeks. Values of zero and 6 weeks were never presented in the delayed or immediate scenarios, respectively. The two runs differed in the type of rewards offered to participants. One of the runs offered participants hypothetical monetary rewards (in USD) to be earned on behalf of either nominee (e.g., $16 for a given

[2] Each continuous trait was discretized into two bins based on high and low endorsement (e.g., extraversion was split into 'extraversion' and 'introversion', conscientiousness was split into 'conscientiousness' and 'directionlessness', etc.). This helped yield items that were highly indicative of both ends of each trait's pole. Five items cross-loaded onto two traits, which is not unexpected given the intercorrelated nature of the Big Five. Participants were only shown the resulting 75 unique items.

target); the other run was comprised of social rewards, offering participants hypothetical time spent with either target (e.g., 16 minutes of time spent with a given close other). The latter condition was explained to participants as added time they could spend with their parent or friend, in addition to the time they already spend with either close other. Values for each type of reward ranged from 2 – 30. There were 49 unique combinations of reward and time pairings (e.g., $2 now versus $18 two weeks from now) for each condition, resulting in 294 unique possible trial types[22]. In the interest of reducing in-scan task demands for participants, each run was comprised of 30 trial types randomly selected from the total 294. It was heavily stressed that participants were to complete the task as if rewards were real, even though they were hypothetical. Participants had 4500 ms to indicate their decision via button box. The trial terminated once they completed their response. Each trial was interspersed with a fixation stimulus '+' and a random jitter (similar parameters as the previous task). Participants completed two runs of the task (1 monetary, 1 social). Each run lasted approximately 90–135 s, with variability owing to the self-paced nature of the task. The visual configuration of task stimuli were generally consistent with prior implementations[22,52]. Both runs were programmed in PsychoPy[53].

Post-Scan Survey. Participants completed a post-scan survey immediately following the scan. During the survey, they provided basic demographic information (sex and age) about their nominated parent and friend, were asked to list approximately 10 adjectives or short phrases describing each close other, answered four written prompts asking them to recall specific memories with each close other, completed a measure of their relationship with each close other, answered questions about how often they see each close other (actual and ideal), and completed a measure of social loss aversion towards each close other (i.e., how upset one would be if they could no longer spend time with a given social partner[54]). Afterwards participants were presented with two one-shot social decision-making questions involving their parent and friend (one with social outcomes at stake, another with monetary outcomes) and finally finished the survey by answering demographic items about themselves. Details about each of these measures can be accessed in the Supplement.

Preprocessing and Statistical Analysis

MRI data preprocessing. fMRIPrep (version 22.0.2)—an open source, freely available tool based on functional neuroimaging software *Nipype*[55]—was used to perform preprocessing MRI preprocessing. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.3.3[56] (RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 6.0.5.1)[57]. Brain surfaces were reconstructed using recon-all (FreeSurfer 7.2.0)[58], and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle[59]. Volume-based spatial normalization to one standard space

(MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both the T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (TemplateFlow ID: MNI152NLin2009cAsym)[60].

For each of the BOLD runs found per subject, the following preprocessing was performed. First, a single-band reference was used as a reference volume. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1)[61]. The BOLD time-series were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration[62]. Co-registration was configured with six degrees of freedom. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions)[63] and Jenkinson (relative root mean square displacement between affines)[61]. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*)[64]. Principal components were estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components were then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al[64]. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask was obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components were dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each[65]. Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries were calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by Patriat et al[66]. The BOLD time-series were resampled into standard space, generating a

preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels. Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Multivariate Pattern Estimation. Multivariate patterns were estimated by modeling imaging data from the preferences task using a General Linear Model (GLM). Each run of the task was modeled using a fixed effects GLM in FSL. Every GLM consisted of the following explanatory variables: A regressor for each of the self, parent, and friend trials where participants rated the extent to which a given item described said target. The duration of each regressor corresponded to the length of the trial (4500 ms). These regressors were convolved with the hemodynamic response function (double gamma). The temporal derivatives of each were also added to account for slice timing effects. Numerous confound regressors were also added to statistically control for noise and head motion. Timeseries from the CSF+WM mask automatically extracted by *fMRIPrep* was added into the model, in addition to the tCompCor component. Finally, the 24 head-motion regressors and individually censored volumes exceeding the FD and DVARS thresholds, both computed by *fMRIPrep*, were added to the GLM. Three linear contrasts were computed ([self - baseline], [parent- baseline], [friend - baseline]).

Representational Similarity Analysis. We used representational similarity analysis [67] to summarize neural representations of others. RSA is useful for determining how various forms of psychological or behavioral information are encoded in the brain, and broadly, for comparing representational structure across modalities. Here we use RSA to examine how information about parents and friends is encoded in the brain, as well as how this information is relevant for social decision-making. The former set of analyses involves conducting various searchlight and ROI analyses, whereas the latter involves extracting individual RDM cells and using them to predict social decision preferences. These analyses are described below in the following three sections.

A Priori ROI Analyses. We implemented RSA on region-of-interest (ROI) data to determine whether there was a 'bias' in the encoding of parent or friend representations in six brain regions, defined *a priori* (mPFC, dmPFC, L NAcc, R NAcc, L TPJ, R TPJ). We did this by constructing an RDM based on between-run distances in multivariate patterns of brain activity. Between-run distances were used to avoid confounds with scanner drift. Concretely, this meant that multivoxel patterns for self, parent, and friend were extracted for a given ROI (using t-statistic images) and distances were computed between patterns from each run for all possible target pairings. Because we were using between run distances, this also meant comparing between run patterns for the same target (e.g., self run 1 was compare to self run 2). We included both (1 – Pearson's *r*) and Euclidean distances as a robustness check—Euclidean

distances are reported in the main document. The mPFC, dmPFC, and TPJ ROIs were defined by locating the respective brain regions on the Neurosynth meta-analytic platform[68], noting coordinates for the global maximum (or within each hemisphere for the TPJ), and drawing a 5 mm radius sphere around the coordinates. The NAcc masks were defined using the Harvard-Oxford subcortical atlas as available in FSL. ROI masks are visualized in the Supplement (Supplementary Figure 1). We then computed pairwise differences between the [self and parent] and [self and friend] RDM cells for each of the six ROIs. Paired differences were modeled as being drawn from a Gaussian distribution such that $Diff_i \sim N(\delta*\sigma, \sigma^2)$. The mean was parameterized as δ*σ so that draws from the distribution are in "native" units but summary statistics of paired differences are in the Cohen's *d* metric (i.e., mean/standard deviation). Prior distributions were assigned to both δ (Cauchy(0, $\frac{1}{\sqrt{2}}$)) and σ (Jeffreys prior, proportional to $\frac{1}{\sigma^2}$). The Bayesian modeling software *stan* (Stan Development Team, 2020) was used to run these analyses via the *rstan* R package (no thinning, 4 chains, 2,000 samples/chain, 1,000 discarded burn-in samples).

Exploratory Searchlight Analyses. We conducted searchlight analyses[69] to identify which brain regions showed the greatest differentiation among the three representations (self, parent, friend). A 125 mm^3 sphere (2 mm radius excluding center voxel $[(2 * 2) + 1]^3$) was moved throughout each participant's brain to compare multivoxel patterns. The first comparison we performed was to directly compare parent and friend representations. This was accomplished by extracting multivariate patterns for the parent and friend t-statistics at each center voxel, vectorizing the voxels within the sphere, and taking the (Pearson) correlation distance (1 – *r*) between the two patterns. As before, we only used between-run comparisons. We used between-run distances within target as a baseline. All between-run distances were placed into a vector and standardized. The mean of within-target distances (e.g., (parent run 1, parent run 2), (friend run 2, friend run 4)) was then subtracted from the mean of (parent, friend) distances. We conducted a second analysis comparing representational overlap between the self and each target. Specifically, distances between neural representations of one's self and one's parent were compared to distances between neural representations of one's self and one's friend in a procedure similar to that described above.

Modeling Social Decision Preferences. We used Bayesian statistics for all regression models involving modeling decision behavior on the social decision-making task. Given that there exists notable intra- and inter-individual variability in decision-making behavior, we chose to model the data at the trial-level, necessitating the use of a hierarchical logistic regression model. The form of the within-person component of the model is notated below.

$$Logit(Dec_{ij}) = b_{0j} + b_{1j}Condition_{ij} + b_{2j}Reward\ Ratio_{ij}$$

This is the core component of the model because it is set-up to encode social decision preferences between parents and friends. Here, the *i*-th decision (*Dec*, 1 = discount, 0 = non-discount) from the *j*-th participant was modeled as a function of a *Condition* dummy code and a *Reward Ratio* variable. *Condition* is coded such that 0 = friend outcome associated with discounting option, parent outcome associated with non-discounting option and 1 = parent

outcome associated with discounting option, friend outcome associated with non-discounting option. The coefficient associated with *Condition* (b_{1j}) encodes social decision preferences between parents and friends because it reflects the mean difference in choice tendencies between the two conditions (after statistically adjusting for other terms in the model). A b_{1j} coefficient > 0 indicates a preference towards parents: there is greater likelihood of choosing to discount when parents are affected by the discounting choice, and a greater likelihood of not discounting when parents are affected by the non-discounting choice. A b_{1j} coefficient < 0 indicates a friend-oriented preference following the same logic. *Reward Ratio* is the quotient of the non-discounting outcome over the discounting outcome and is included here to statistically adjust for varying levels of rewards across trials as well as providing a 'sanity check' for participant engagement (greater reward ratio should generally be associated with a reduced likelihood of discounting). All three terms in the within-person component of the model were allowed to vary randomly across participants.

Allowing the b_{1j} coefficient to vary randomly across participants meant that we could model variability in individual decision preferences. We did so by running separate models, entering different between-participant variables each time. These between-participant variables can be conceptualized as being allowed to interact with the *Condition* dummy code (as a cross-level interaction in a one-level model). The following sets of between-person predictors were tested (each group in a separate model): (i) dissimilarity values for self and parent, as well as self and friend calculated from behavioral ratings, (ii-v) neural dissimilarities from each ROI (bilateral NAcc and TPJ ROIs were entered in the same model). Linguistic preference scores were entered as part of a three-way interaction between neural dissimilarity and *Condition* in subsequent analyses. All continuous predictors were grand mean centered prior to running the model. All models were implemented using the *brms* package in the R statistical software library (no thinning, 8 chains, 2500 samples/chain, 1250 discarded burn-in samples). A weakly informative standard normal prior was placed on all slope coefficients and brms default priors were used for all other model parameters.

Because the output of Bayesian analyses is a distribution, we performed statistical inference in a graded manner and thus do not report nor use p-values to conduct frequentist inference. We summarized posterior distributions using the mean of the relevant statistic (e.g., Cohen's d, regression coefficient) and an 89% highest density credible interval (HDI). We used 89% credible intervals upon the recommendation that wider intervals (e.g., 95%) are more sensitive to Monte Carlo sampling error[70,71]. We used two criteria to perform inferences: (i) examining whether the HDI contained zero, and (ii) Kruschke's Region of Practical Equivalence Method (ROPE)[72,73]. ROPE values for all analyses ranged from -0.1 to 0.1. We judged evidence to be robust if the HDI did not include zero and the HDI fell outside of ROPE, moderate if the HDI did not include zero but overlapped with the ROPE, and modest if the HDI overlapped with the ROPE but not span both the entire region [74]. We considered HDIs that fell entirely within the ROPE to be evidence for a null effect, and HDIs that spanned the ROPE (i.e., the HDI poked out of both ends of the ROPE) to be an inconclusive effect. A more detailed explanation of these inferential criteria can be accessed in the Supplement.

Study 2.

Datasets. Study 2 was comprised of secondary analyses of three large existing datasets. We used one large dataset (Dataset 1) to construct and preliminarily validate a linguistic signature of social decision preferences. The written text data and social decision items in this dataset were collected as part of a broader study aiming to answer an orthogonal research question[54]. Two additional datasets (Dataset 2, Dataset 3) were sourced from previously published reports on social decision-making by the first author[1,21,22] to further validate the linguistic signature.

Dataset 1 was comprised of five independently collected subsamples (S1.1 – S1.5). Participants in each subsample were asked to nominate a parent and close friend of their choice and provide approximately ten adjective or short phrases (e.g., 'happy go lucky', 'pain in the butt') describing each close other. Participants in S1.1 also prompted to recall four types of memories involving each other – an instance in which they felt supported by the close other, in which supported their close other, in which their close other let them down, and their favorite memory with their close other. Social decision preferences were assessed with a one-shot dictator game question in which individuals allocated a hypothetical $100 pot between parents and friends in $5 increments, and another in which they indicated who they would rather spend time with given a day with no other obligations or commitments. Subsamples S1.1 – S1.5 totaled N = 870 participants (range of subsample size n = 45 – 454). Subsamples were recruited from the UCLA undergraduate psychology subject pool and online crowdsourcing platforms (Mturk, Prolific) in 2022.

Dataset 2 was comprised of five independently collected subsamples (S2.1 – S2.5). Participants in each subsample were asked to nominate a parent and close friend of their choice, provide a memory they share with each target, and list several words and phrases describing the target. Social decision preferences were assessed with a version of the Columbia Card Task[75]—a multi-trial risk-taking paradigm—that was modified to capture social decision preferences when making choices with conflicting outcomes between a pair of others[21]. Subsamples S2.1 – S2.5 totaled N = 574 participants (range of subsample size n = 46 – 225). Subsamples were recruited from the UCLA undergraduate psychology subject pool and the broader West Los Angeles community between 2016 - 2020.

Dataset 3 was comprised of four subsamples (S3.1 – S3.4). Some of the participants in these subsamples also provided data for subsamples in Datasets 1 and 2. The distinctive contribution of this dataset relative to the others is that it uses a different social decision-making paradigm. As with the other datasets, participants were prompted to nominate a parent and close friend. Some subsamples followed the text generation procedure described for Dataset 1, others follow the procedure for Dataset 2, meaning that all subsamples provided adjectives and short phrases describing each other, and some also provided a memory of each other. Social decision preferences were assessed with a version of a multi-trial delay discounting task that was modified to capture social decision preferences when making choices with conflicting outcomes between a pair of others[22]. Subsamples S3.1 – S3.4 totaled N = 920 participants (range of subsample size n = 61 – 454). Subsamples were recruited from the UCLA undergraduate psychology subject pool (2018-2019) and from an online crowdsourcing platform (2022).

Further information about each dataset, such as participant demographics and methodological details, can be accessed in the original publications cited above. Participants in

all datasets were told they could not nominate family members and current romantic partners as friends. Self-reported relationship quality between participants and their nominated parents and friends was available in all datasets. All participants provided consent in accordance with the policies of the UCLA Institutional Review Board. Depending on their method of recruitment, participants were either compensated in course credit or monetary remuneration (USD). All study materials and code are publicly available on the Open Science Framework (OSF; osf.io/fzds6). For privacy reasons, we are unable to share the raw or preprocessed text data.

Constructing the Linguistic Signature via Machine Learning. Since our goal was to examine whether linguistic representations of close others encode information about social decision preferences, we trained a model to predict social decisions from semantic features of written text data. The first step in this process was to use cross-validation on the subsamples from Dataset 1 to settle on model specifications (representation definition, type of model, hyperparameters). The pipeline for this task was comprised of the following steps: (i) social decision criterion selection, (ii) text preprocessing, (iii) representation definition, (iv) leave-one-out cross-validation using each subsample, and (v) estimating model weights on the entire dataset using the model with the best cross-validation accuracy.

Defining Social Decision Preferences for Signature Construction. We used the forced choice question about which close other the participant would rather spend a free day with (coded such that parent = 1, friend = 0) as our measure of social decision preferences. More details about this item and our rationale can be accessed in the Supplement. Crucially, choice preferences were not particularly imbalanced across the five subsamples (S1.1 = 51% chose parents over friends; S1.2 = 42%; S1.3 = 42%; S1.4 = 46%; S1.5 = 42%)

Text Preprocessing. Written text data was preprocessed in a manner consistent with other similar studies [32], implemented with the TextBlob python library. The following preprocessing pipeline was applied to each participant's data. We first removed stopwords, as defined by TextBlob, in addition to the terms 'favorite' and 'memory' (as many participants responses to prompts asking to recall a memory began with this phrase). Afterwards, each word was forced into its lowercase form and lemmatized. Words were then tagged as part of speech and filtered. Only nouns, adjectives, verbs and adverbs were in the preprocessed word list, corresponding to the following TextBlob tags: 'JJ', 'JJR', 'JJS', 'NN', 'NNS', 'NNP', 'NNPS', 'RB', 'RBR', 'RBS', 'VB', 'VBD', 'VBG', 'VBP', 'VBZ'. Parent and friend memories were preprocessed separately, meaning that each participant had a set of a preprocessed words from written text data about their nominated parent, and another preprocessed word set about their friend.

Representation Definition. We used word embeddings to define linguistic representations of parents and friends from participants' written text data. This approach extracted semantic meaning from the preprocessed word set in a bottom-up fashion via the use of pretrained Word2Vec word embeddings (300 dimensions from the Google news dataset). Embeddings within a word set were averaged and then reduced via principal components analysis (PCA) and used in subsequent modeling. This approach is described in greater detail in the Supplement.

Leave-One-Dataset-Out Cross-Validation. We used leave-one-dataset-out cross-validation to determine the best specifications for defining a linguistic signature of social decision preferences. We specifically sought to identify the best type of machine learning model, set of hyperparameters, and representation definition. We identified three possible types of machine learning models: a support vector classifier (SVC), a ridge regression classifier (RRC), and regularized logistic regression (RLR). Hyperparameters for each model were selected using a grid search within each iteration of cross-validation. The candidate hyperparameters for all possible hyperparameters (C, γ, α) were {1, 5, 10, 50, 100}.

Our cross-validation approach took advantage of the independent subsamples in Dataset 1—we chose to leave-one-dataset out for each iteration of cross-validation. Since the goal of cross-validation is to estimate out-of-sample modeling performance, we reasoned the natural partitioning of each subsample as independent datasets best served this goal. This meant that each iteration of cross-validation proceeded such that one subsample was scaled and set aside, the other four were aggregated and scaled, a grid search was performed to find the best hyperparameter(s), the best hyperparameter was then used to estimate model weights on the four aggregated subsamples, and the model weights were finally applied to the left-out subsample to estimate the predictive accuracy of the model. This process was repeated five times, leaving one of the subsamples out each time, for all of the three models described above.

Estimating Model Weights. As previously noted, the binary forced choice question about spending time with either a parent (coded as 1) or friend (coded as 0) served as our criterion. The principal components derived from both the parent and friend text data were entered into the model as predictors (60 total for the Word2Vec representation, 30 total for the LIWC representation). Thus, depending on how we defined the linguistic representation, there were either 60 or 30 model weights. Together, the weights indicate how a 'preference' towards one close other is embedded in the linguistic representation.

Validating the Linguistic Signature. If our linguistic signature is truly capturing social decision preferences between parents and friends, then the signature should generalize to novel measures and predict preferences in new datasets. To that end, we used the weights from our signature to compute preference scores (the dot product between model weights and principal components from processed text data) for novel data and then tested whether those preference scores could predict social decision preferences on previously untested social decision tasks. Here, we used decisions from a one-shot dictator game where participants had to directly allocate money between parents and friend (Dataset 1), a decision-making task that pitted gains and losses for parents against those for friends in a risky context (Dataset 2), and a decision-making task where participants had to choose to allocate money for, or time spent with, their parents and friends in a delay discounting context (Dataset 3; the same task used in the Study 1). Finally, because social decision preferences are often predicted by attitudes towards social partners, we also tested whether the linguistic signature could predict relationship quality scores for parents and friends (available in all datasets). These validation tests are described in greater detail below. Bayesian statistics were used for all analyses. The inferential criteria from Study 1 were also applied to Study 2 analyses.

Social Preferences in the One-Shot Dictator Game. The one-shot dictator game, collected in Dataset 1, asked participants to allocate a hypothetical pot of $100 between their nominated parent and friend. Values varied by increments of $5, ranging from [$100 parent, $0 friend] to [$0 parent, $100 friend], resulting in a 21-point ordinal variable. Responses were recoded to be the percentage of money allocated towards the parent (0, .05, .10 … .90, .95, 100). A robust Pearson correlation between preference scores and dictator game allocations was computed using the Bayesian software package stan (implemented in the R statistical software platform). We note that while the text data for this validation step is not novel, the outcome variable (dictator game allocations) is, and it is valuable to know whether preference scores can predict decision preferences in different contexts within-participants. More details about the dictator game can be accessed in the Supplement.

Social Preferences in a Risky Context. Linguistic preference scores were next validated by determining whether they could predict parent vs friend preferences in a risky decision context. Participants in this dataset (Dataset 2) were required to make choices between a risky and a safe outcome. The risky outcome carried the potential for either a hypothetical monetary gain or a hypothetical monetary loss according to a known probability; the safe outcome was neutral (no gain or loss). The interests of parents and friends were pitted against each other such that any potential gain from a risky choice on a given trial would be credited to one partner, whereas any potential loss would be credited to the other partner. Participants completed multiple trials across two runs of the task, one in which risky gains benefitted one partner (e.g., parent) while the losses were incurred by the other partner (e.g., friend), and another run in which the opposite was true. Hierarchical Bayesian logistic regression was used to model the trial-level likelihood of making a risky (vs safe) decision as a function of a parameter that encoded choice preferences on the task and other trial-level covariates. We refer readers to Guassi Moreira et al.[21] for details about the form of the within-person component of the model. We entered preference scores derived from the linguistic signature as a between-person predictor to the model, allowing it to interact with the choice preference parameter from the model. Because the choice preference parameter was allowed to vary randomly across subjects, the ensuing interaction term codes for how choice preferences on the task vary as a function of preference scores derived from the linguistic signature. The brms package was used to fit the models (4 chains, 1000 discarded burn-in samples, 2000 samples/chain, no thinning).

Social Preferences in a Delay Discounting Context. Linguistic preference scores were further validated by testing whether they could predict parent vs friend preferences in a delay discounting context. This is the same task used in Study 1, with the only difference being that the full set of trials were used. Hierarchical Bayesian logistic regression was used to model the trial-level likelihood of making a discounting decision (i.e., choosing the immediate over delayed reward) as a function a parameter that encoded choice preferences on the task and another trial-level covariate (same within-person model as Study 1). We ran separate models for choices with monetary and social outcomes. Preference scores derived from the linguistic signature were entered into this model in the same manner described above. The brms package was also used to fit the models (same specifications listed above).

Predicting Relationship Quality Using the Linguistic Signature. Last, we opted to relate preference scores derived from the linguistic signature to self-reported relationship quality with parents and friends. In doing so, our goal was to subject the linguistic signature to another validatory step that involved predicting the *attitudes* that often drive social decision behavior. A meaningful result for this test would indicate that motivationally relevant information about one's attitude towards a close other is encoded in their linguistic representation. Data for this test come from all three datasets used here, totaling 1,641 participants. Information about the self-report measure used here (Inventory of Parent and Peer Attachment) and its reliability is included in the Supplement.

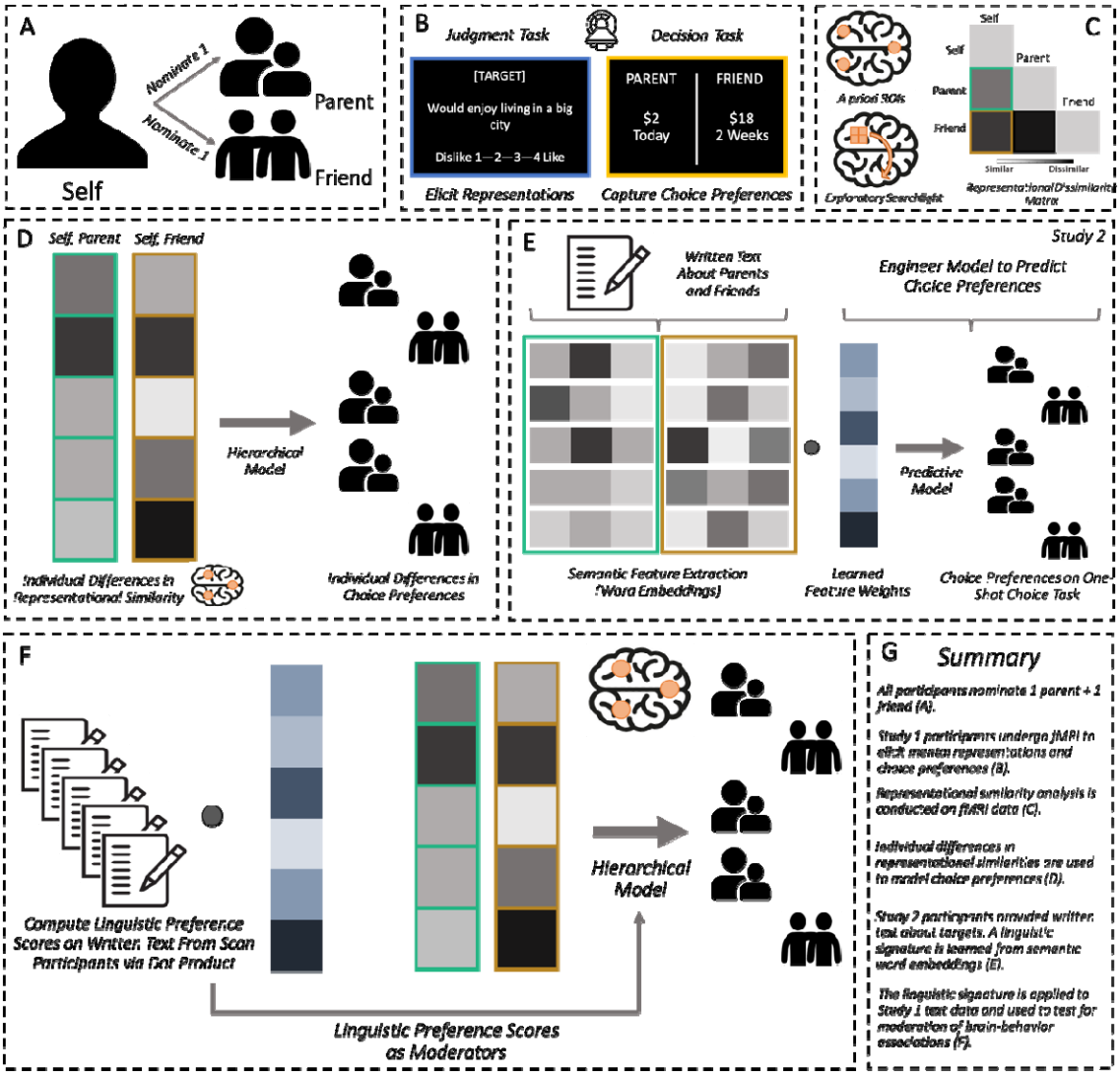
Unpacking Associations between Linguistic Preference Scores and Neural Representations. To unpack the statistical relationship between neural representations and linguistic preference scores we conducted an exploratory, descriptive analysis using the most recent version of the Linguistic Inquiry and Word Count (LIWC-22) software[76] to identify the semantic themes that drive linguistic preference scores and then correlated them with neural representations. We did this in two steps.

First, we passed the written text data of a subset of Study 2 participants (n = 1,641) through LIWC-22 and extracted the prevalence of the following thematic categories (individual facets in parentheses): core summary themes (clout, analytical thinking, authentic, emotional tone), drives (affiliation, achievement, power), cognition (all-or-none, cognitive processes, insight, causation, discrepancy, tentative, certitude, differentiation), affect (positive tone, negative tone, positive emotion, negative emotion, anxiety, anger, sadness), social processes (prosocial behavior, politeness, interpersonal conflict, moralization, communication), culture (politics, ethnicity, technology), lifestyle (leisure, home, work, money, religion), states (mental health, need, want, acquire, lack, fulfilled, fatigue), and motives (reward, risk, curiosity, allure). The prevalence of each of these themes in parent and friend text was separately correlated with linguistic preference scores. We kept themes that showed a correlation of $r > |.05|$. This threshold was chosen because it roughly corresponded to the threshold for statistical significance in this subsample and because it minimized the risk of falsely excluding an interesting or meaningful theme. The surviving themes and correlations are reported in Table 5.

Second, we passed the text data from Study 1 participants through LIWC-22 to extract the prevalence of the themes identified in Table 5, and then correlated them with the most relevant neural ROIs identified in Figures 4-5. These correlations are depicted as Cleveland plots in Figures 6-7. The most consistently strong themes for parents and friends were identified by taking the magnitude (absolute value) of each theme's correlation with neural data and averaging it across the ROIs. These are listed in the top right corner of Figures 6-7. Because the purpose of this analysis was meant to descriptively unpack what is happening in our sample and help generate future research, we did not conduct inferential statistics on the estimates.

Acknowledgments
Preparation of this research was supported by a National Science Foundation SBE Postdoctoral Research Fellowship to JFGM (award number 2104629) and the UCLA Department of Psychology (CP). We thank Dr. Rick Dale for comments on the linguistic analyses. We thank the members of the Computational Social Neuroscience Laboratory at UCLA for their feedback on study concept and manuscript.

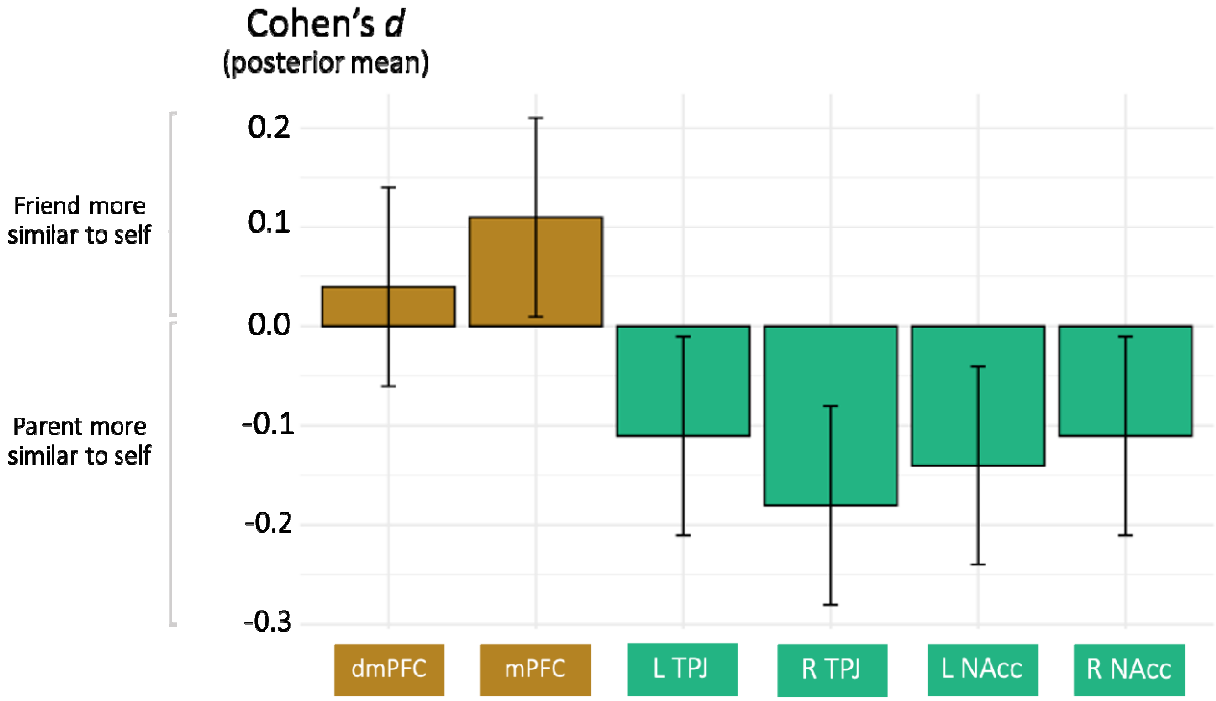
Figure 1. Study Overview



Note. (*A*) Participants in both studies nominated a parent and same-age friend of their choice who was not a current romantic partner or family member. (*B*) Participants in Study 1 underwent fMRI scanning while completing a task designed to elicit mental representations of the self, parent, and friend, as well as a social decision-making task that involved allocating monetary and social rewards between parent and friend. *(C)* RSA was used to compare parent and friend representations (directly, and relative to the self) in six *a priori* ROIs and in exploratory searchlight analyses. *(D)* Individual differences in self - other (parent and friend) representational similarity from the ROIs were used to predict decision preferences between parents and friends. *(E)* Study 2 participants provided written memories about, and/or adjectives describing their nominated parent and friend and completed a one-shot social decision task. Semantic features for parent and friend text were extracted (Word2Vec embeddings) and used to create and validate a linguistic signature of participants' decision preferences between parents and friends on the decision task. The linguistic signature yielded 'linguistic preference scores' that quantified implicit social choice preferences embedding in linguistic representations. These scores generalized to additional data (other social decision tasks, relationship

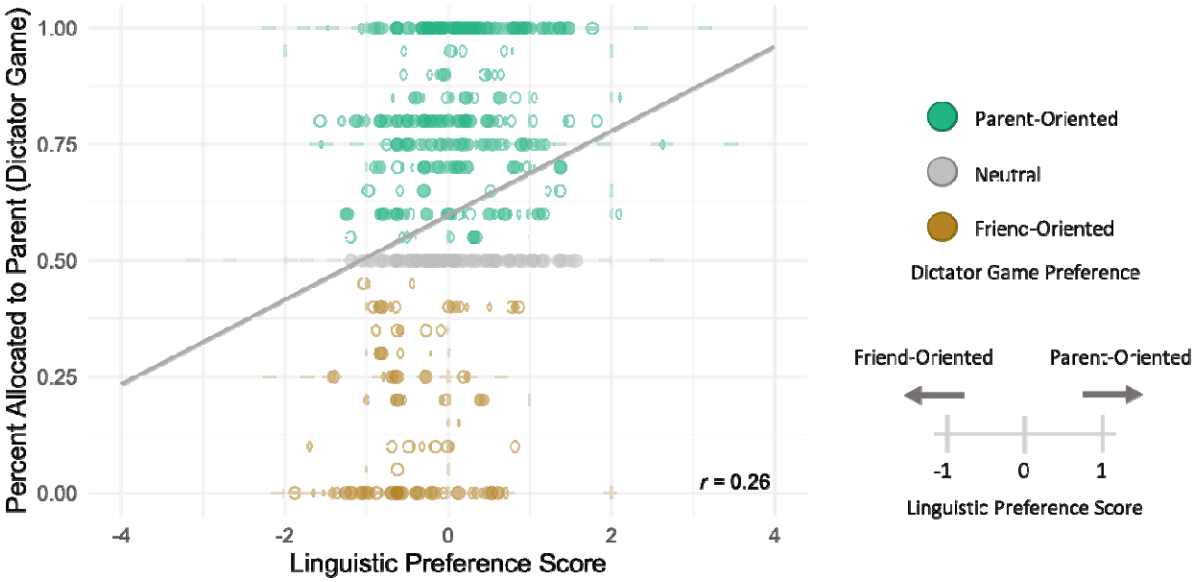
characteristic such quality, social loss aversion). *(F*) The linguistic signature was used to compute linguistic preference scores on written text data obtained from Study 1 participants. Scores were entered as moderators in models predicting decision preferences from neural representational similarity. The study workflow is summarized in (*G*).

Figure 2. Brain regions involved in social cognition and value-based processes differentially construct representations of parents and friends in relation to the self.



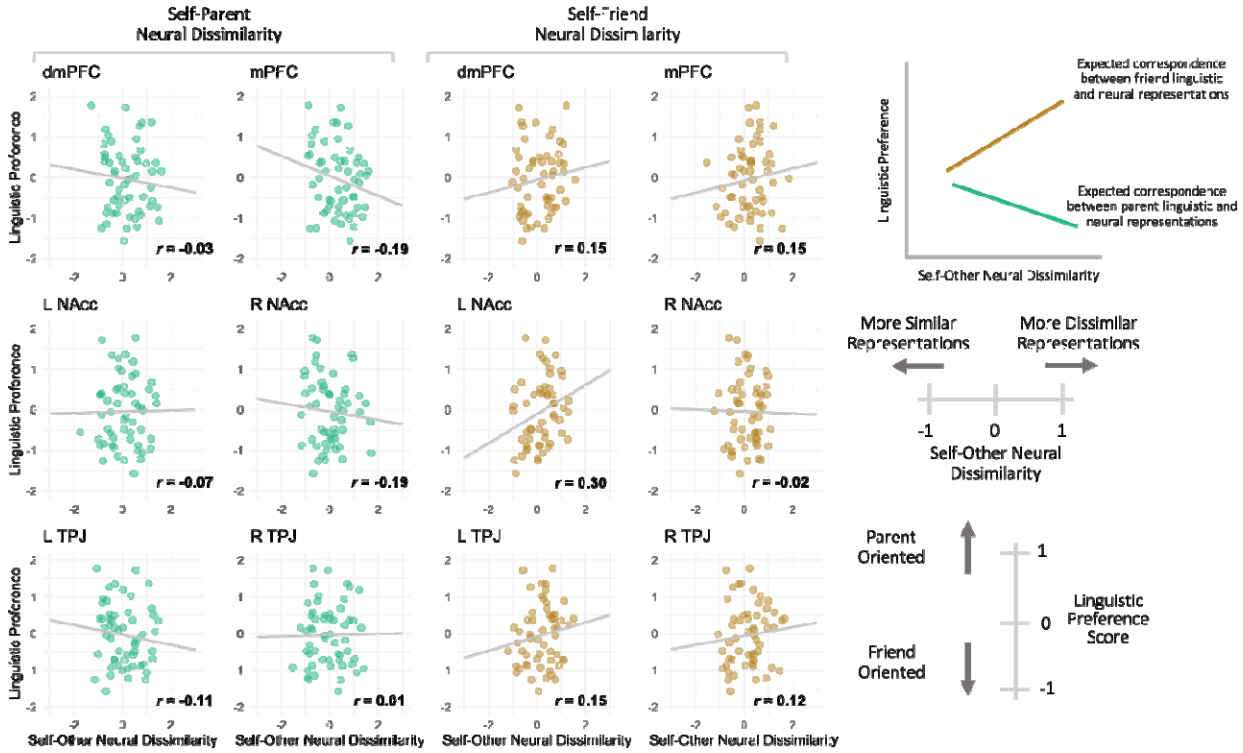
Note. 'dmPFC' refers to dorsomedial prefrontal cortex. 'mPFC' refers to medial prefrontal cortex. 'TPJ' refers to temporoparietal junction. 'NAcc' refers to nucleus accumbens. 'L' refers to left and 'R' refers to right. Cohen's *d* values reflect paired differences between (self, parent) - (self, friend) dissimilarities of multivoxel response patterns extracted from the six brain regions displayed here. Error bars correspond to posterior standard deviations.

Figure 3. Linguistic preference scores directly predict social decision preferences on a novel decision task (dictator game).



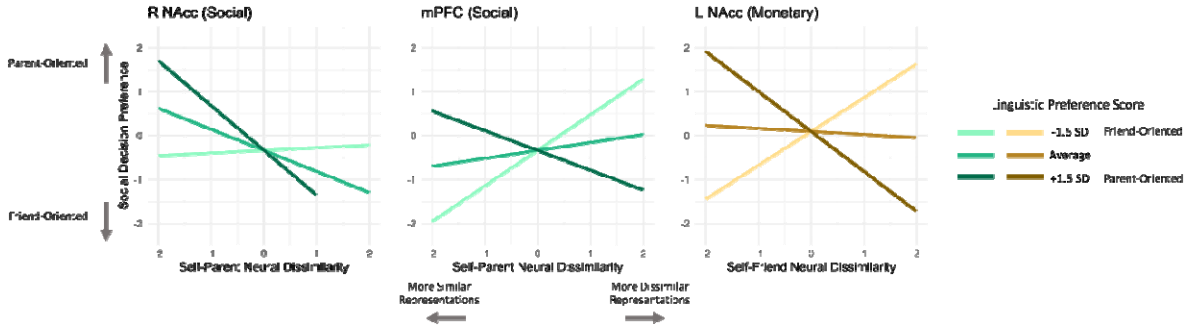
Note. Linguistic preference scores refer to model implied preferences between a parent and friend based on written text data; relatively greater values indicate a parent-over-friend preference, relatively lower values indicate a friend-over-parent preference. The Y-axis depicts the percentage of a hypothetical $100 endowment allocated to one's parent (rather than one's friend); greater values thus indicate stronger parent-over-friend preferences in this task, whereas lower values indicate the opposite. Linguistic preference scores were calculated with weights trained on the Word2Vec-based linguistic representations.

Figure 4. Neural and linguistic representations of parents and friends are correlated in the TPJ, NAcc, and dmPFC.



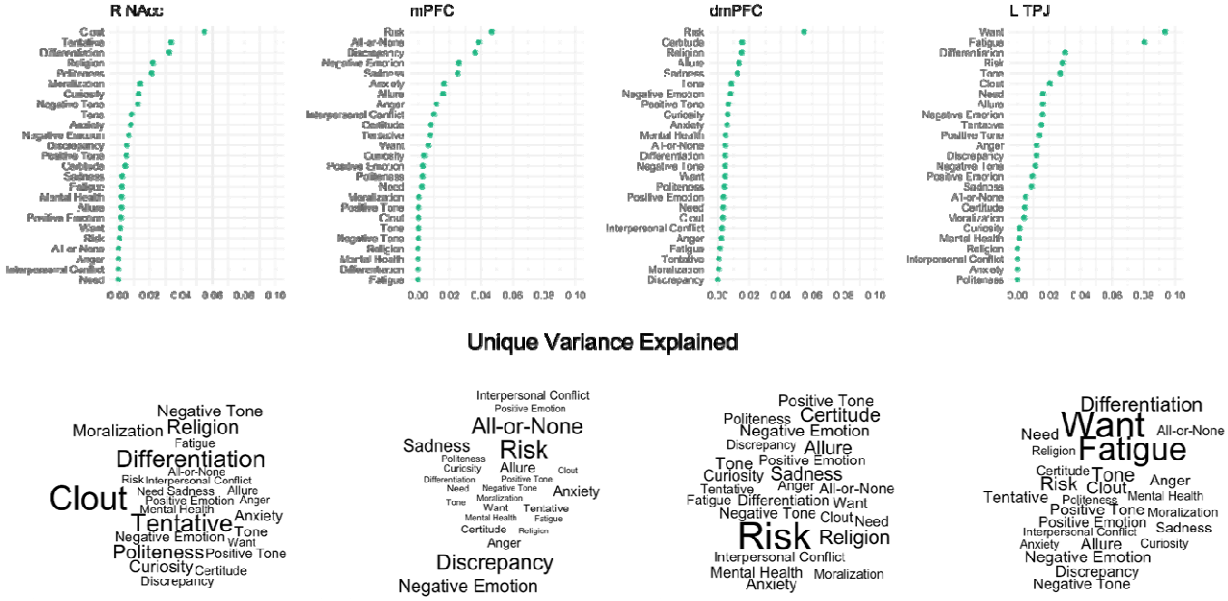
Note. Linguistic preference scores refer to model implied social decision preferences between a parent and friend based on written text data; relatively greater values indicate a parent-over-friend preference, relatively lower values indicate a friend-over-parent preference. The Y-axis depicts neural dissimilarity between self and other (parent, friend) representations across six neural ROIs. Linguistic preference scores were calculated with weights trained on Word2Vec-based linguistic representations.

Figure 5. Correspondence between linguistic and neural representations in three ROIs is associated with stronger decision preferences



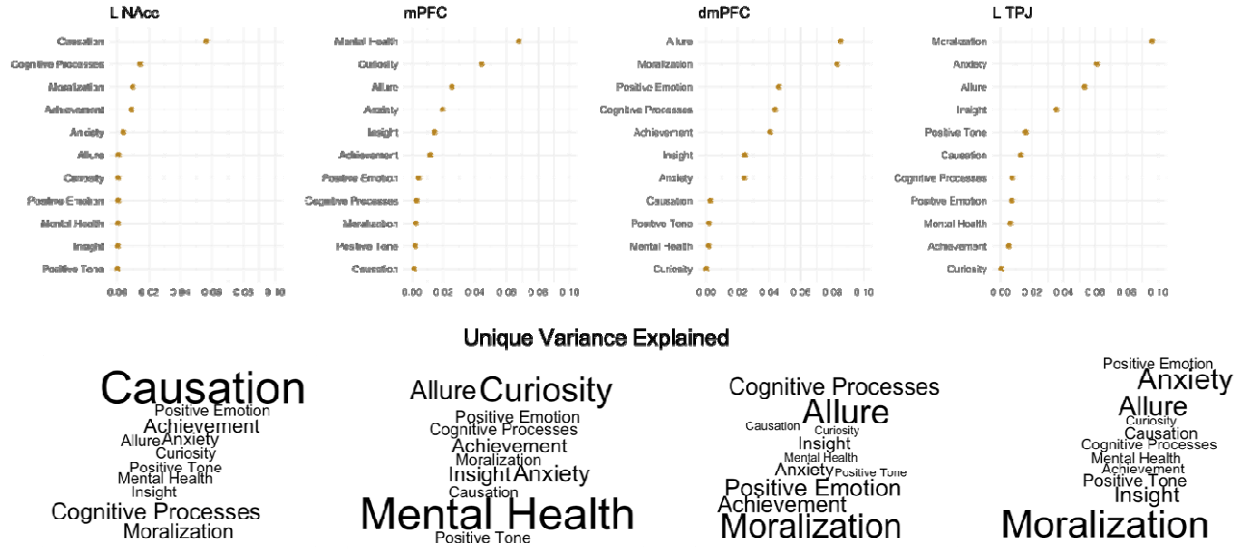
Note. Linguistic preference scores refer to model implied social decision preferences between a parent and friend based on written text data; relatively greater values indicate a parent-over-friend preference, relatively lower values indicate a friend-over-parent preference. The Y-axis represents choice preferences between parent and friend. The X-axis depicts neural representational dissimilarity between the self and a target other (parent, friend) for a given ROI. The relevant ROI is listed at the top of the top of plot, along with the outcome domain of the decision task (monetary or social). Linguistic preference scores were calculated with weights trained on the Word2Vec-based linguistic representations.

Figure 6. Unpacking the correspondence between linguistic and neural representations of parents: Unique variance explained in self-parent neural dissimilarity by each linguistic theme.



Note. Linguistic themes analyzed here are those from the ‘parent’ column of Table 5. ROIs analyzed here are those showing the strongest relationship with linguistic preference scores (Figures 4-5). Datapoints represent the squared semi-partial correlation between the prevalence of a linguistic theme in written text data (about parents) and self-parent neural dissimilarity in an ROI (i.e., the unique variance each theme accounts for in the neural data given all other themes). This is an exploratory descriptive analysis meant to understand what semantic content is encoded in self-other neural representations. Pearson’s *r* coefficient was used. Each term in the word cloud corresponds to a linguistic theme, and their sizes are scaled by the unique variance explained.

Figure 7. Unpacking the correspondence between linguistic and neural representations of friends: Unique variance explained in self-friend neural dissimilarity by each linguistic theme.



Note. Linguistic themes analyzed here are those from the ‘friend’ column of Table 5. ROIs analyzed here are those showing the strongest relationship with linguistic preference scores (Figures 4-5). Datapoints represent the squared semi-partial correlation between the prevalence of a linguistic theme in written text data (about friends) and self-friend neural dissimilarity in an ROI (i.e., the unique variance each theme accounts for in the neural data given all other themes). This is an exploratory descriptive analysis meant to understand what semantic content is encoded in self-other neural representations. Pearson’s *r* coefficient was used. Each term in the word cloud corresponds to a linguistic theme, and their sizes are scaled by the unique variance explained.

Table 1. Friend representations are more similar to self representations in cortical midline regions, whereas parent representations are more similar to self representations in the TPJ and NAcc

ROI	Paired Difference [89% HDI]
dmPFC	0.04 [-0.15, 0.25]
mPFC	*0.11 [-0.10, 0.30]*
L TPJ	*-0.11 [-0.31, 0.09]*
R TPJ	*-0.18 [-0.38, 0.03]*
L NAcc	*-0.14 [-0.35, 0.08]*
R NAcc	*-0.11 [-0.30, 0.10]*

Note. ‘dmPFC’ refers to dorsomedial prefrontal cortex. ‘mPFC’ refers to medial prefrontal cortex. ‘TPJ’ refers to temporoparietal junction. ‘NAcc’ refers to nucleus accumbens. ‘L’ rand ‘R’ refer to left and right hemisphere, respectively. Values in the table reflect Cohen’s *d* statistics (posterior mean) for paired comparisons between dissimilarities of multivoxel representations of self and parent, and self and friend, in each brain region. Neural dissimilarities were calculated using Euclidean distance. Values in brackets reflect 89% highest density credible intervals drawn around the posterior distribution. ROPE was determined as [-0.1, 0.1]. Difference where the HDI partially overlaps with ROPE (moderate evidence) are italicized.

Table 2. Modeling social decision preferences as a function of the overlap of neural representations of self and parent, and self and friend, in dmPFC and mPFC

	Monetary Outcomes		Social Outcomes	
	dmPFC	mPFC	dmPFC	mPFC
Intercept	-2.49 [-3.05, -1.93]	-2.52 [-3.07, -1.95]	-0.92 [-1.33, -0.50]	-0.92 [-1.32, -0.50]
Condition	-0.08 [-0.57, 0.44]	-0.06 [-0.63, 0.48]	-0.27 [-0.81, 0.21]	-0.28 [-0.78, 0.19]
Reward ratio	-2.96 [-3.56, -2.39]	-2.94 [-3.52, -2.34]	-1.29 [-1.60, -0.93]	-1.28 [-1.62, -0.96]
Self-parent neural dissimilarity	0.45 [0.01, 0.90]	0.31 [-0.14, 0.80]	0.02 [-0.34, 0.38]	0.09 [-0.28, 0.45]
Self-friend neural dissimilarity	-0.22 [-0.65, 0.23]	0.17 [-0.31, 0.62]	0.01 [-0.36, 0.37]	0.33 [-0.04, 0.66]
Condition x Self-parent neural dissimilarity	***-0.98 [-1.51, -0.44]***	-0.26 [-0.81, 0.34]	-0.23 [-0.78, 0.28]	-0.17 [-0.70, 0.36]
Condition x Self-friend neural dissimilarity	-0.07 [-0.58, 0.45]	-0.24 [-0.82, 0.35]	-0.28 [-0.79, 0.26]	*-0.58 [-1.11, -0.07]*

Note. Coefficients are on a logit scale and reflect means from a posterior distribution. 'Condition' indicates when parent or friend outcomes were associated with the discounting or delay options and encodes social decision preferences (positive values indicate a parent-over-friend preference, negative values indicate a friend-over-parent preference). Reward ratio reflects the division of the non-discounting option over the discounting option. Self-parent and self-friend neural dissimilarities refer to the distance between multivoxel neural patterns for oneself and a parent or friend, extracted from the brain region listed in the column heading. Neural dissimilarities were calculated via Euclidean distance. Values in brackets reflect 89% highest density credible intervals drawn around the posterior distribution. Interactions reflect associations between neural dissimilarities and social decision preferences. Coefficients where the HDI partially overlaps with ROPE (modest evidence) are italicized; coefficients where the HDI does not overlap with ROPE (strong evidence) are bolded and italicized.

Table 3. Modeling social decision preferences as a function of the overlap of neural representations of self and parent, and self and friend, bilaterally in the TPJ and NAcc

	Monetary Outcomes		Social Outcomes	
	TPJ	NAcc	TPJ	NAcc
Intercept	-2.51 [-3.01, -1.99]	-2.56 [-3.07, -1.95]	-0.91 [-1.32, -0.51]	-0.94 [-1.35, -0.51]
Condition	-0.06 [-0.59, 0.50]	-0.04 [-0.61, 0.49]	-0.28 [-0.80, 0.20]	-0.28 [-0.78, 0.22]
Reward ratio	-2.90 [-3.46, -2.30]	-3.02 [-3.62, -2.43]	-1.28 [-1.59, -0.94]	-1.31 [-1.65, -0.99]
Self-parent neural dissimilarity (L)	0.19 [-0.23, 0.62]	0.07 [-0.44, 0.57]	-0.28 [-0.63, 0.09]	-0.07 [-0.47, 0.29]
Self-friend neural dissimilarity (L)	-0.49 [-0.93, -0.05]	0.09 [-0.40, 0.60]	-0.42 [-0.79, -0.03]	-0.30 [-0.68, 0.08]
Self-parent neural dissimilarity (R)	0.62 [0.20, 1.07]	-0.13 [-0.60, 0.35]	0.15 [-0.24, 0.53]	0.19 [-0.19, 0.54]
Self-friend neural dissimilarity (R)	0.70 [0.25, 1.12]	-0.08 [-0.52, 0.38]	0.44 [0.07, 0.81]	-0.02 [-0.37, 0.32]
Condition x Self-parent neural dissimilarity (L)	-0.25 [-0.84, 0.36]	-0.39 [-0.97, 0.21]	-0.14 [-0.68, 0.42]	0.05 [-0.48, 0.62]
Condition x Self-friend neural dissimilarity (L)	0.03 [-0.56, 0.66]	*-0.48 [-1.11, 0.08]*	-0.23 [-0.80, 0.36]	-0.01 [-0.59, 0.54]
Condition x Self-parent neural dissimilarity (R)	-0.46 [-1.05, 0.13]	0.20 [-0.38, 0.74]	0.41 [-0.17, 0.99]	*-0.50 [-1.08, 0.03]*
Condition x Self-friend neural dissimilarity (R)	-0.42 [-1.01, 0.20]	***0.81 [0.23, 1.35]***	0.03 [-0.57, 0.56]	0.02 [-0.51, 0.54]

Note. Coefficients are on a logit scale and reflect means from a posterior distribution. 'Condition' indicates when parent or friend outcomes were associated with the discounting or delay options and thus encodes social decision preferences (positive values indicate a parent-over-friend preference, negative values indicate a friend-over-parent preference). Reward ratio reflects the

division of the non-discounting option over the discounting option. Self-parent and self-friend neural dissimilarities refers to the distance between multivoxel neural patterns for oneself and a parent or friend, extracted from the brain region listed in the column heading. Neural dissimilarities were calculated via Euclidean distance. Values in brackets reflect 89% highest density credible intervals drawn around the posterior distribution ‘L’ rand ‘R’ refer to left and right hemisphere, respectively. Interactions reflect associations between neural dissimilarities and social decision preferences. Interaction slopes where the HDI partially overlaps with ROPE (modest evidence) are italicized; interaction slopes where the HDI does not overlap with ROPE (strong evidence) are bolded and italicized.

Table 4. Correlations between linguistic preference scores and neural representational dissimilarities between self and other across two close others (targets) and six ROIs

ROI	Target	Robust Correlation [89% HDI]
dmPFC	Parent	-0.03 [-0.24, 0.18]
dmPFC	Friend	*0.15 [-0.07, 0.35]*
mPFC	Parent	*-0.19 [-0.41, 0.00]*
mPFC	Friend	*0.15 [-0.05, 0.36]*
L NAcc	Parent	-0.07 [-0.27, 0.14]
L NAcc	Friend	***0.30 [0.11, 0.51]***
R NAcc	Parent	*-0.19 [-0.38, 0.04]*
R NAcc	Friend	-0.02 [-0.24, 0.20]
L TPJ	Parent	*-0.11 [-0.32, 0.10]*
L TPJ	Friend	*0.15 [-0.04, 0.37]*
R TPJ	Parent	0.01 [-0.21, 0.22]
R TPJ	Friend	*0.12 [-0.08, 0.33]*

Note. 'dmPFC' refers to dorsomedial prefrontal cortex. 'mPFC' refers to medial prefrontal cortex. 'TPJ' refers to temporoparietal junction. 'NAcc' refers to nucleus accumbens. 'L' and 'R' refer to left and right hemisphere, respectively. Values in brackets reflect 89% highest density credible intervals drawn around the posterior distribution. Correlation values reflect a robust version of Pearson's *r* obtained by using *t*-distributed likelihood function. Neural dissimilarities were obtained with Euclidean distance. Linguistic preference scores were derived from Word2Vec-based linguistic representations. Correlations where the HDI partially overlaps with ROPE (modest evidence) are italicized; correlations where the HDI does not overlap with ROPE are bolded and italicized (robust evidence). Negative correlation values are expected for parent, positive values are expected for friend.

Table 5. Linguistic themes from parent and friend linguistic representations most strongly correlated with linguistic preference scores

Thematic Category	Parent text and linguistic preference score correlation	Friend text and linguistic preference score correlation
Core		
Clout	0.05	-
Tone	0.15	-
Drives		
Achievement	-	-0.06
Cognition		
All-or-None	0.06	-
Cognitive Processes	-	-0.11
Insight	-	-0.09
Causation	-	-0.09
Discrepancy	0.07	-
Tentative	0.05	-
Certitude	0.09	-
Differentiation	0.05	-
Affect		
Positive Tone	0.14	0.07
Negative Tone	-0.24	-
Positive Emotion	0.09	0.05
Negative Emotion	-0.24	-
Anxiety	-0.21	-0.05
Anger	-0.12	-
Sadness	-0.11	-
Social Processes		
Politeness	0.11	-
Interpersonal Conflict	-0.05	-
Moralization	0.05	0.08
Lifestyle		
Religion	0.07	-
States		
Mental Health	-0.13	0.05
Need	-0.05	-
Want	0.05	-
Fatigue	-0.09	-
Motives		
Risk	-0.05	-
Curiosity	-0.06	-0.12
Allure	0.17	0.08

Note. Linguistic themes were obtained by independently passing parent and friend written representations through LIWC-22 and then correlating prevalence for each theme with linguistic preference scores. Positive correlation values in the ‘parent’ column indicate that greater prevalence of a theme is associated with a higher linguistic preference score and thus a stronger parent-over-friend social decision preference. Negative correlation values in the ‘friend’ column indicate that greater prevalence of a theme is associated with a lower linguistic

preference score and thus a stronger friend-over-parent social decision preference. This is an exploratory descriptive analysis meant to understand the semantic content that drives linguistic preference scores. Correlations below $r = |0.05|$ were excluded. This threshold was chosen to be inclusive given the exploratory nature of the analysis.

References

1. Guassi Moreira, J. F., Tashjian, S. M., Galván, A. & Silvers, J. A. Is social decision making for close others consistent across domains and within individuals? *J. Exp. Psychol. Gen.* **149**, 1509–1526 (2020).

2. FeldmanHall, O. & Chang, L. J. Social Learning. in *Goal-Directed Decision Making* 309–330 (Elsevier, 2018). doi:10.1016/B978-0-12-812098-9.00014-0.

3. Yu, H., Siegel, J. Z. & Crockett, M. J. Modeling Morality in 3-D: Decision-Making, Judgment, and Inference. *Top. Cogn. Sci.* **11**, 409–432 (2019).

4. Kao, C.-H., Feng, G. W., Hur, J. K., Jarvis, H. & Rutledge, R. B. Computational models of subjective feelings in psychiatry. *Neurosci. Biobehav. Rev.* **145**, 105008 (2023).

5. Austerweil, J. L. *et al.* The Impact of Other-Regarding Preferences in a Collection of Non-Zero-Sum Grid Games. in (The AAAI Press, 2016).

6. Crockett, M. J., Siegel, J. Z., Kurth-Nelson, Z., Dayan, P. & Dolan, R. J. Moral transgressions corrupt neural representations of value. *Nat. Neurosci.* **20**, 879–885 (2017).

7. Cole, T. & Bruno Teboul, Jc. Non-zero-sum collaboration, reciprocity, and the preference for similarity: Developing an adaptive model of close relational functioning. *Pers Relatsh* **11**, 135–160 (2004).

8. Sampaio, W. M., Freitas, A. L., Rêgo, G. G., Morello, L. Y. N. & Boggio, P. S. Effects of co-players' identity and reputation in the public goods game. *Sci. Rep.* **13**, 13520 (2023).

9. Hackel, L. M., Zaki, J. & Van Bavel, J. J. Social identity shapes social valuation: evidence from prosocial behavior and vicarious reward. *Soc. Cogn. Affect. Neurosci.* **12**, 1219–1228 (2017).

10. Huettel, S. A. & Kranton, R. E. Identity economics and the brain: uncovering the mechanisms of social conflict. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 680–691 (2012).

11. Broom, T. W., Chavez, R. S. & Wagner, D. D. Becoming the King in the North: identification with fictional characters is associated with greater self-other neural overlap. *Soc. Cogn. Affect. Neurosci.* **16**, 541–551 (2021).

12. Hassabis, D. *et al.* Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cereb. Cortex* **24**, 1979–1987 (2014).

13. Wang, Y. *et al.* Dynamic neural architecture for social knowledge retrieval. *Proc Natl Acad Sci USA* **114**, E3305–E3314 (2017).

14. Mitchell, J. P., Banaji, M. R. & Macrae, C. N. The link between social cognition and self-referential thought in the medial prefrontal cortex. *J. Cogn. Neurosci.* **17**, 1306–1315 (2005).

15. Chavez, R. S. & Wagner, D. D. The neural representation of self is recapitulated in the brains of friends: A round-robin fMRI study. *J. Pers. Soc. Psychol.* **118**, 407–416 (2020).

16. Mitchell, J. P., Macrae, C. N. & Banaji, M. R. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* **50**, 655–663 (2006).

17. Fleming, A. C. & Slank, K. L. Making a choice: Self-other differences in decision making in risky situations. *N Am J Psychol* **17**, 633–648 (2015).

18. Ho, M. K., Cushman, F., Littman, M. L. & Austerweil, J. L. Communication in action: Planning and interpreting communicative demonstrations. *J. Exp. Psychol. Gen.* **150**, 2246–2272 (2021).

19. Sarathy, V., Scheutz, M., Kenett, Y., Allaham, M. M. & Malle, B. F. Mental Representations and Computational Modeling of Context-Specific Human Norm Systems. (2017).

20. Hawkins, R., Liu, I., Goldberg, A. & Griffiths, T. Respect the code: Speakers expect novel conventions to generalize within but not across social group boundaries. *Proceedings of the Annual Meeting of the Cognitive Science Society* (2021).

21. Guassi Moreira, J. F., Tashjian, S. M., Galván, A. & Silvers, J. A. Parents versus peers: assessing the impact of social agents on decision making in young adults. *Psychol. Sci.* **29**, 1526–1539 (2018).

22. Guassi Moreira, J. F., Tashjian, S. M., Galván, A. & Silvers, J. A. Computational and motivational mechanisms of human social decision making involving close others. *J. Exp. Soc. Psychol.* **93**, (2021).

23. Crone, E. A. & Fuligni, A. J. Self and others in adolescence. *Annu. Rev. Psychol.* **71**, 447–469 (2020).

24. Crosnoe, R. & Johnson, M. K. Research on Adolescence in the Twenty-First Century. *Annu. Rev. Sociol.* **37**, 439–460 (2011).

25. Greenberg, M. T., Siegel, J. M. & Leitch, C. J. The nature and importance of attachment relationships to parents and peers during adolescence. *J. Youth Adolesc.* **12**, 373–386 (1983).

26. Syed, M. & Mitchell, L. L. Race, ethnicity, and emerging adulthood. *Emerging Adulthood* **1**, 83–95 (2013).

27. Fuligni, A. J. The need to contribute during adolescence. *Perspect. Psychol. Sci.* **14**, 331–343 (2019).

28. Tamir, D. I. & Mitchell, J. P. Anchoring and adjustment during social inferences. *J. Exp. Psychol. Gen.* **142**, 151–162 (2013).

29. Saxe, R. & Powell, L. J. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol. Sci.* **17**, 692–699 (2006).

30. Pfeifer, J. H. & Peake, S. J. Self-development: integrating cognitive, socioemotional, and neuroimaging perspectives. *Dev. Cogn. Neurosci.* **2**, 55–69 (2012).

31. Delgado, M. R. *et al.* Viewpoints: Dialogues on the functional role of the ventromedial prefrontal cortex. *Nat. Neurosci.* **19**, 1545–1552 (2016).

32. Fatima, A., Li, Y., Hills, T. T. & Stella, M. Dasentimental: detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning. *BDCC* **5**, 77 (2021).

33. Cadinu, M. R. & Rothbart, M. Self-anchoring and differentiation processes in the minimal group setting. *J. Pers. Soc. Psychol.* **70**, 661–677 (1996).

34. Krienen, F. M., Tu, P.-C. & Buckner, R. L. Clan mentality: evidence that the medial prefrontal cortex responds to close others. *J. Neurosci.* **30**, 13906–13915 (2010).

35. Todd, A. R. & Tamir, D. I. Factors that amplify and attenuate egocentric mentalizing. *Nat. Rev. Psychol.* **3**, 164–180 (2024).

36. Froehlich, L., Dorrough, A. R., Glöckner, A. & Stürmer, S. Similarity Predicts Cross-National Social Preferences. *Soc. Psychol. Personal. Sci.* **12**, 1486–1498 (2021).

37. Fischer, I. & Savranevski, L. The effect of similarity perceptions on human cooperation and confrontation. *Sci. Rep.* **13**, 19849 (2023).

38. Guassi Moreira, J. F. *et al.* Value-based neural representations predict social decision preferences. *Cereb. Cortex* **33**, 8605–8619 (2023).

39. Tamir, D. I. & Thornton, M. A. Modeling the predictive social mind. *Trends Cogn Sci (Regul Ed)* **22**, 201–212 (2018).

40. Lin, C. & Thornton, M. Evidence for bidirectional causation between trait and mental state inferences. *J. Exp. Soc. Psychol.* **108**, 104495 (2023).

41. Kourtis, D., Vandemaele, P. & Vingerhoets, G. Concurrent Cortical Representations of Function- and Size-Related Object Affordances: An fMRI Study. *Cogn. Affect. Behav. Neurosci.* **18**, 1221–1232 (2018).

42. McMahon, E. & Isik, L. Seeing social interactions. *Trends Cogn Sci (Regul Ed)* **27**, 1165–1179 (2023).

43. de Wit, M. M., de Vries, S., van der Kamp, J. & Withagen, R. Affordances and neuroscience: Steps towards a successful marriage. *Neurosci. Biobehav. Rev.* **80**, 622–629 (2017).

44. Karan, M. *et al.* Giving to others and neural processing during adolescence. *Dev. Cogn. Neurosci.* **56**, 101128 (2022).

45. Delton, A. W. *et al.* Cognitive foundations for helping and harming others: Making welfare tradeoffs in industrialized and small-scale societies. *Evolution and Human Behavior* **44**, 485–501 (2023).

46. Sznycer, D. Value computation in humans. *Evolution and Human Behavior* **43**, 367–380 (2022).

47. Brandner, P., Güroğlu, B., van de Groep, S., Spaans, J. P. & Crone, E. A. Happy for Us not Them: Differences in neural activation in a vicarious reward task between family and strangers during adolescent development. *Dev. Cogn. Neurosci.* **51**, 100985 (2021).

48. Thornton, M. A. & Mitchell, J. P. Theories of person perception predict patterns of neural activity during mentalizing. *Cereb. Cortex* **28**, 3505–3520 (2018).

49. John, O. P. & Srivastava, S. *The big-five trait taxonomy: History, measurement, and theoretical perspectives*. vol. 2 102–138 (1999).

50. De Raad, B. *The Big Five Personality Factors: The psycholexical approach to personality*. (Hogrefe & Huber Publishers, 2000).

51. de Leeuw, J. R., Gilbert, R. A. & Luchterhandt, B. jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments. *JOSS* **8**, 5351 (2023).

52. Seaman, K. L. *et al.* Adult age differences in decision making across domains: Increased discounting of social and health-related rewards. *Psychol. Aging* **31**, 737–746 (2016).

53. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).

54. Guassi Moreira, J. F. & Parkinson, C. A Behavioral Signature for Quantifying the Social Value of Interpersonal Relationships with Specific Others. *PsyArXiv* (2023).

55. Gorgolewski, K. J. *et al.* Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinformatics* **5**, 13 (2011).

56. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).

57. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).

58. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

59. Klein, A. *et al.* Mindboggling morphometry of human brains. *PLoS Comput. Biol.* **13**, e1005350 (2017).

60. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R. & Collins, D. L. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* **47**, S102 (2009).

61. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).

62. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).

63. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).

64. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).

65. Satterthwaite, T. D. *et al.* An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).

66. Patriat, R., Reynolds, R. C. & Birn, R. M. An improved model of motion-related signal changes in fMRI. *Neuroimage* **144**, 74–82 (2017).

67. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).

68. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).

69. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc Natl Acad Sci USA* **103**, 3863–3868 (2006).

70. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. (Chapman and Hall/CRC, 2015). doi:10.1201/9781315372495.

71. Makowski, D., Ben-Shachar, M. & Lüdecke, D. bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *JOSS* **4**, 1541 (2019).

72. Kruschke, J. K. Bayesian estimation supersedes the t test. *J. Exp. Psychol. Gen.* **142**, 573–603 (2013).

73. Kruschke, J. K. Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspect. Psychol. Sci.* **6**, 299–312 (2011).

74. Wiens, B. L. Choosing an equivalence limit for noninferiority or equivalence studies. *Control. Clin. Trials* **23**, 2–14 (2002).

75. Figner, B., Mackinlay, R. J., Wilkening, F. & Weber, E. U. Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **35**, 709–730 (2009).

76. Boyd, R. L., Ashokkumar, A., Seraj, S. & Pennebaker, J. W. The Development and Psychometric Properties of LIWC-22. *Unpublished* (2022) doi:10.13140/rg.2.2.23890.43205.

77. Lebeau, G. *et al.* Delay discounting of gains and losses, glycemic control and therapeutic adherence in type 2 diabetes. *Behav. Processes* **132**, 42–48 (2016).

78. Golsteyn, B. H. H., Grönqvist, H. & Lindahl, L. Adolescent time preferences predict lifetime outcomes. *The Economic Journal* **124**, F739–F761 (2014).

79. Armsden, G. C. & Greenberg, M. T. The inventory of parent and peer attachment: individual differences and their relationship to psychological well-being in adolescence. *J. Youth Adolesc.* **16**, 427–454 (1987).

80. Spiegelhalter, D. J., Freedman, L. S. & Parmar, M. K. B. Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **157**, 357 (1994).

81. Kruschke, J. K. Rejecting or accepting parameter values in bayesian estimation. *Advances in Methods and Practices in Psychological Science* **1**, 251524591877130 (2018).

82. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. *arXiv* (2013) doi:10.48550/arxiv.1301.3781.

Supplement

Study 1 Participant Demographics. The full sample was comprised of 46 individuals assigned the female sex at birth (73%). In terms of gender, 43 individuals identified as women (68.3%), 17 as men (27.0%), 2 as non-binary (3.2%), and 1 (1.6%) declined to respond. Ethnically, 12 identified as Hispanic/Latinx (19.0%). Racially, 24 identified as Asian/American (38.1%), 23 identified as Caucasian (36.5%), 7 identified as mixed race/multi-ethnic (11.1%), 5 identified as another category ('Other', 7.9%), 1 declined to respond (1.6%), and 0 identified as Native Hawaiian/Pacific Islander and American Indian/Alaska Native (0%). In terms of sexual orientation, 46 individuals identified as straight/heterosexual (73%), 8 identified as bisexual (12.7%), 4 identified as homosexual (gay/lesbian, 6.3%), 3 identified as another category ('Other', 4.8%), and 2 identified as asexual (3.2%). Finally, 57 participants reported relying on their parents for financial assistance (90.5%). The self-reported median household income among participants was $100,000 (range = $18,000 - $3,000,000).

Study 1 Parent-Close Friend Nomination. Just prior to receiving training for the fMRI tasks, participants were told the purpose of the study was to understand how the brain encodes information about close others and how it uses said information when making choices about others. Participants were thus asked to nominate a parent and their closest friend. The closest friend was not allowed to be a current romantic partner or a family member. The majority of participants chose their mother (84%) whereas the remainder chose their father (mean age of nominated parent: 51.4 years, SD = 5.6 years). The majority (69%) of participants nominated a close friend assigned the female sex at birth (mean age of nominated parent: 19.1 years, SD = 1.2 years).

Rationale for Choosing a Delay Discounting Task. Our rationale for delay discounting to examine social decision preferences is threefold. First, discounting decisions are both pervasive in everyday life and are thought to be important for shaping life adjustment outcomes[77,78]. Second, computerized discounting tasks deployed in research settings are flexible in their configuration, allowing researchers to study social decision behavior in the face of different reward outcomes [52]. Last, we previously used discounting tasks to reliably tap social decision preferences involving specific targets [22], thereby providing the current study with a baseline with which to perform basic manipulation checks.

Rationale for Using Forced Choice Time Spent Question in Constructing the Linguistic Signature. We chose this metric for several reasons. The binary nature of the item makes for an intuitive classification metric with which to evaluate model performance (percent accuracy). Relatedly, a binary forced choice prevents participants from a neutral scale response bias, and requires them to disclose an unambiguous preference. Finally, framing social decision preferences in terms of spending time (a finite resource subject to opportunity cost) with others is externally valid and likely generalizes better to less externally common decision scenarios (e.g., allocating money between others). Because the model is based on preferences that mimic a real-world scenario, it will arguably generalize better to artificial scenarios employed in lab studies (as opposed to the opposite).

Description of Post-Scan Survey Measures.

The following is an extended description of the each of the post-scan survey measures listed in the main document. Not all post-scan survey measures were analyzed.

Relationship Quality. Relationship quality with each close other was assessed using the Inventory of Parent and Peer Attachment[79]. Participants were instructed to use a 5-point Likert scale (1 = *almost never or never*, 5 = *almost always or always*) to answer 28 items about their relationship with their parent and a second set of 25 items about their relationship with their friend. Participants were instructed to complete the items with their nominated parent and close friend in mind, and the measure's wording was lightly modified according to this prompt (sample parent item: "My parent respects my feelings"; sample friend item: "When we discuss things, my friend considers my point of view"). Responses were reverse-scored where appropriate and averaged to create a single composite score for relationship quality with each respective other. Reliabilities for each sample ranged from good to excellent and have been reported elsewhere[1,21,22].

One-Shot Dictator Game. Participants completed a one-trial dictator game (Fehr & Camerer, 2007) involving their parent and friend. Dictator games are traditionally used in behavioral economics and social psychology to study fairness, altruism, and reciprocity; we use them as a vehicle for assessing social preferences between targets in a monetary outcome domain. Participants were presented with the following prompt: "*Suppose you had $100 to split between the parent you nominated and your closest friend. Use the dropdown box below to indicate how you would allocate the money.*" Allocations ranged from '$100 parent, $0 friend' to '$0 parent, $100 friend' in $5 increments, totaling 21 possible response options. As noted, participants used a drop-down menu to indicate their allocation for the game. Responses were recoded to be the percentage of money allocated to the parent.

Forced Choice Question About Spending Time. Participants completed a one-shot forced choice question about spending time between their parent and friend. Participants were presented with the following prompt. "*Suppose you had a free afternoon with no obligations or commitments. Assuming you could only spend it with one person, who would you spend it with the parent you nominated, or your closest friend?"*. Participants then chose between their parent and friend. Responses were binary coded such that 1 = parent, 0 = friend.

Social Loss Aversion. Participants completed a measure of social loss aversion[54] regarding their parent and friend. Social loss aversion is intended to measure one's attitude towards the prospect of no longer being able to spend time with a particular individual. The measure is administered a single-item questionnaire prompt asking participants how upset they would be if they could no longer spend time with a given individual, capturing responses with a 100-point sliding scale (0 = "not at all", 100 = "extremely"). The greater the value indicated on the sliding scale, the greater the participant's aversion to the social loss of that given individual.

Actual and Ideal Time Spent with Social Targets. Participants indicated how much actual ("For an average month, how many days do you see [TARGET]") and ideal ("Ideally, for an average month, how many days would you see [TARGET]") time they spend with each of their nominated targets via a drop-down menu.

Inferential Criteria for Bayesian Analyses.

In this study, we evaluated the weight of statistical evidence using a combination the Region of Practical Equivalence (ROPE) method[72,73] and noninferiority testing[74,80]. Traditionally, the ROPE method entails specifying a symmetric region in parameter space around the null value, calculating a credible interval (CI), and then evaluating whether the CI intersects at all with ROPE. CIs that do not overlap with ROPE result in a rejection of the null hypothesis, CIs that are entirely contained within ROPE result in acceptance of the null, and any other configuration falls in undecided. Originating in clinical trials research, noninferiority testing is similar: if the CI partly overlaps with the upper limit of ROPE but does not pass through below the lower limit (regardless of its relation to the null value), the parameter value is declared to be *noninferior* to the null value (the same method can be used in the other direction, e.g., for 'nonsuperior' inference).

Here we combine these methods in the following manner. If the CI does not overlap with ROPE, we deem the strength of evidence to be robust. If the CI does not overlap with the null value (usually zero), but partly overlaps with ROPE, we deem the strength of evidence to be moderate. If the CI overlaps with ROPE and the null value (in the manner usually described in noninferiority testing), we deem the strength of evidence to be modest. If the CI completely *spans* ROPE (i.e., overlaps and exceeds both ends of ROPE), deem the evidence to be inconclusive. If the CI is contained within rope, we interpret the finding to support the null result. We deliberately follow these inferential criteria because it displaces the inferential focus from a problematic binary 'significant vs not-significant' label[81] to a more nuanced conclusion describing the *degree* of evidence in relation to meaningful benchmarks. We report standardized statistics and thus use a ROPE of [-0.1, 0.1] for all Bayesian analyses, unless noted elsewhere. We report 89% highest density credible intervals for all analyses.

Defining Linguistic Representations Using Word2Vec Embeddings. We extracted word embedding vectors for each word in the parent and friend word sets separately, and then averaged the vectors of all set words together. We used pretrained embeddings from Google's 300-dimension Word2Vec model [82], as implemented in the genism python package. Words in the word set that did not have a corresponding word embedding vector were excluded from the representation. In this approach, the semantic content of each participant's text data for a given close other (parent, friend) was summarized as a composite 300-dimensional vector. We used principal components analysis (PCA) to reduce dimensionality in order to aid modeling. Both parent and friend composite vectors for each subject were reduced to 30 dimensions.

Using LIWC to Unpack the Relationship Between Linguistic Preference Scores and Neural Representations. We used LIWC to characterize written text data of participants' nominated parent and friend. The following 46 thematic dimensions were selected given their theoretical and ecological relevance. Four from LIWC's core summary dimensions: analytical thinking, clout, authentic, emotional tone. Three from the 'Drives' category: Affiliation,

achievement, and power. Nine from the ‘Cognition’ category: All-or-none, cognitive processes, insight, causation, discrepancy, tentative, certitude, differentiation, and memory. Seven from the ‘Affect’ category: Positive tone, negative tone, positive emotion, negative emotion, anxiety, anger, and sadness. Six from the “Social Processes” category: Social behavior, prosocial behavior, politeness, interpersonal conflict, moralization, and communication. Six from the “Lifestyle” category: Lifestyle, leisure, home, work, money, and religion. Six from the “States” category: Need, want, acquire, lack, fulfilled, and fatigue. Four from the “Motives” category: Reward, risk, curiosity, and allure. Finally, the mental health dimension was added from the “Health” category.

Summary variables indicating the prevalence of category words for each dimension were extracted from parent and friend memories in the Study 2 dataset and correlated with linguistic preference scores. We thresholded these correlations at $r = |0.05|$ to find themes to correlate with neural representational data in the Study 1 dataset. Given the exploratory nature of the study, we chose the relatively low threshold to minimize false negatives. The surviving linguistic themes were used to compute semi-partial correlations with self-other neural representational overlap in the brain regions identified in Figure 5. Squaring these correlations yields the unique percent variance that each theme accounts for in the representational information for a given brain region after partialing out the effect of all other themes.

Supplementary Table 1. Paired differences between self-parent representational overlap and self-friend representational overlap across six ROIs computed using Pearson's *r*

ROI	Paired Difference [89% HDI]
dmPFC	-0.10 [-0.31, 0.09]
mPFC	0.03 [-0.19, 0.22]
L TPJ	-0.09 [-0.29, 0.13]
R TPJ	-0.13 [-0.32, 0.07]
L NAcc	-0.05 [-0.27, 0.15]
R NAcc	-0.06 [-0.26, 0.15]

Note. 'dmPFC' refers to dorsomedial prefrontal cortex. 'mPFC' refers to medial prefrontal cortex. 'TPJ' refers to temporoparietal junction. 'NAcc' refers to nucleus accumbens. 'L' rand 'R' refer to left and right hemisphere, respectively. Values in the table reflect Cohen's *d* statistics (posterior mean) for paired comparisons between dissimilarities of multivoxel representations of self and parent, and self and friend in each ROI. Values in brackets reflect 89% highest density credible intervals drawn around the posterior distribution. ROPE was determined as [-0.1, 0.1]. Difference where the HDI partially overlaps with ROPE (moderate evidence) are italicized. Neural dissimilarities were computed with 1 – Pearson's *r* (results using Euclidean distances to compare neural patterns are provided in the main text).

Supplementary Table 2. Modeling social decision preferences as a function of the overlap of neural representations of self and parent, and self and friend, in dmPFC and mPFC (computed using Pearson's *r*)

	Monetary Outcomes		Social Outcomes	
	dmPFC	mPFC	dmPFC	mPFC
Intercept	-2.51 [-3.08, -1.95]	-2.52 [-3.08, -1.95]	-0.93 [-1.37, -0.52]	-0.92 [-1.33, -0.50]
Condition	-0.08 [-0.62, 0.44]	-0.06 [-0.61, 0.49]	-0.28 [-0.76, 0.21]	-0.28 [-0.77, 0.19]
Reward Ratio	-2.95 [-3.56, -2.38]	-2.94 [-3.54, -2.39]	-1.29 [-1.61, -0.95]	-1.28 [-1.63, -0.97]
Self-parent neural dissimilarity	0.19 [-0.23, 0.62]	0.31 [-0.65, 0.20]	0.14 [-0.19, 0.48]	0.09 [-0.57, 0.10]
Self-friend neural dissimilarity	-0.28 [-0.72, 0.14]	0.17 [-0.86, -0.01]	0.04 [-0.29, 0.42]	0.33 [-0.34, 0.36]
Condition x Self-parent neural dissimilarity	-0.31 [-0.84, 0.22]	-0.26 [-0.35, 0.75]	*-0.44 [-0.93, 0.06]*	-0.17 [-0.50, 0.49]
Condition x Self-friend neural dissimilarity	*0.46 [-0.08, 0.99]*	-0.24 [-0.40, 0.69]	-0.35 [-0.90, 0.13]	*-0.58 [-1.01, 0.02]*

Note. Coefficients are on a logit scale and reflect means from a posterior distribution. 'Condition' indicates when parent or friend outcomes were associated with the discounting or delay options and encodes social decision preferences (positive values indicate a parent-over-friend preference, negative values indicate a friend-over-parent preference). Reward ratio reflects the division of the non-discounting option over the discounting option. Self-parent and self-friend neural dissimilarities refer to the distance between multivoxel neural patterns for oneself and a parent or friend, extracted from the brain region listed in the column heading. Neural dissimilarities were calculated via 1 - Pearson's *r* (results using Euclidean distances to compare neural patterns are provided in the main text). Values in brackets reflect 89% highest density credible intervals drawn around the posterior distribution. Interactions reflect associations between neural dissimilarities and social decision preferences. Coefficients where the HDI partially overlaps with ROPE (modest evidence) are italicized; coefficients where the HDI does not overlap with ROPE (strong evidence) are bolded and italicized.

Supplementary Table 3. Correlations between linguistic preference scores and neural representational dissimilarities between self and other (computed using Pearson's *r*) across two close others (targets) and six ROIs

ROI	Target	Robust Correlation [89% HDI]
dmPFC	Parent	-0.07 [-0.28, 0.15]
dmPFC	Friend	0.08 [-0.13, 0.28]
mPFC	Parent	*-0.16 [-0.36, 0.04]*
mPFC	Friend	0.10 [-0.12, 0.32]
L NAcc	Parent	0.11 [-0.11, 0.34]
L NAcc	Friend	*0.14 [-0.08, 0.36]*
R NAcc	Parent	*-0.14 [-0.37, 0.04]*
R NAcc	Friend	0.09 [-0.13, 0.30]
L TPJ	Parent	*-0.20 [-0.41, 0.00]*
L TPJ	Friend	0.03 [-0.17, 0.25]
R TPJ	Parent	0.05 [-0.16, 0.27]
R TPJ	Friend	0.06 [-0.15, 0.28]

Note. 'dmPFC' refers to dorsomedial prefrontal cortex. 'mPFC' refers to medial prefrontal cortex. 'TPJ' refers to temporoparietal junction. 'NAcc' refers to nucleus accumbens. 'L' and 'R' refer to left and right hemisphere, respectively. Values in brackets reflect 89% highest density credible intervals drawn around the posterior distribution. Correlation values reflect a robust version of Pearson's *r* obtained by using *t*-distributed likelihood function. Neural dissimilarities were obtained with 1 – Pearson's *r* (results using Euclidean distances to compare neural patterns are provided in the main text). Linguistic preference scores were derived from Word2Vec-based linguistic representations. Correlations where the HDI partially overlaps with ROPE (modest evidence) are italicized; correlations where the HDI does not overlap with ROPE are bolded and italicized (robust evidence). Negative correlation values are expected for parent, positive values are expected for friend.

Supplementary Table 4. Predicting social decision preferences with interactions between linguistic preference scores and neural dissimilarity scores (dmPFC, mPFC)

Term	dmPFC – Monetary	mPFC – Monetary	dmPFC – Social	mPFC – Social
Intercept	-2.50 [-3.07, -1.93]	-2.44 [-3.03, -1.88]	-0.94 [-1.40, -0.52]	-0.89 [-1.33, -0.47]
Condition	-0.03 [-0.55, 0.51]	-0.15 [-0.74, 0.38]	-0.28 [-0.78, 0.25]	-0.34 [-0.84, 0.16]
Reward ratio	-2.96 [-3.56, -2.40]	-2.95 [-3.52, -2.35]	-1.30 [-1.64, -0.97]	-1.29 [-1.61, -0.96]
Self-parent neural dissimilarity	0.47 [0.00, 0.96]	0.28 [-0.22, 0.76]	0.03 [-0.35, 0.41]	0.04 [-0.33, 0.43]
Self-friend neural dissimilarity	-0.20 [-0.66, 0.28]	0.09 [-0.39, 0.56]	0.05 [-0.35, 0.44]	0.31 [-0.07, 0.68]
Linguistic preference score	0.15 [-0.27, 0.61]	0.21 [-0.25, 0.67]	-0.05 [-0.40, 0.30]	-0.07 [-0.42, 0.29]
Condition x Self-parent neural dissimilarity	-1.07 [-1.62, -0.51]	-0.30 [-0.87, 0.31]	-0.20 [-0.74, 0.37]	-0.05 [-0.59, 0.49]
Condition x Self-friend neural dissimilarity	-0.05 [-0.60, 0.48]	-0.11 [-0.70, 0.47]	-0.38 [-0.95, 0.20]	-0.60 [-1.17, -0.09]
Condition x Linguistic preference score	-0.50 [-1.01, 0.01]	-0.53 [-1.06, 0.05]	0.19 [-0.33, 0.69]	0.12 [-0.42, 0.62]
Self-parent neural dissimilarity x Linguistic preference score	0.11 [-0.40, 0.62]	0.18 [-0.35, 0.73]	0.16 [-0.23, 0.55]	0.02 [-0.39, 0.43]
Self-friend neural dissimilarity x Linguistic preference score	0.08 [-0.47, 0.63]	-0.30 [-0.70, 0.47]	0.12 [-0.33, 0.58]	-0.23 [-0.65, 0.21]

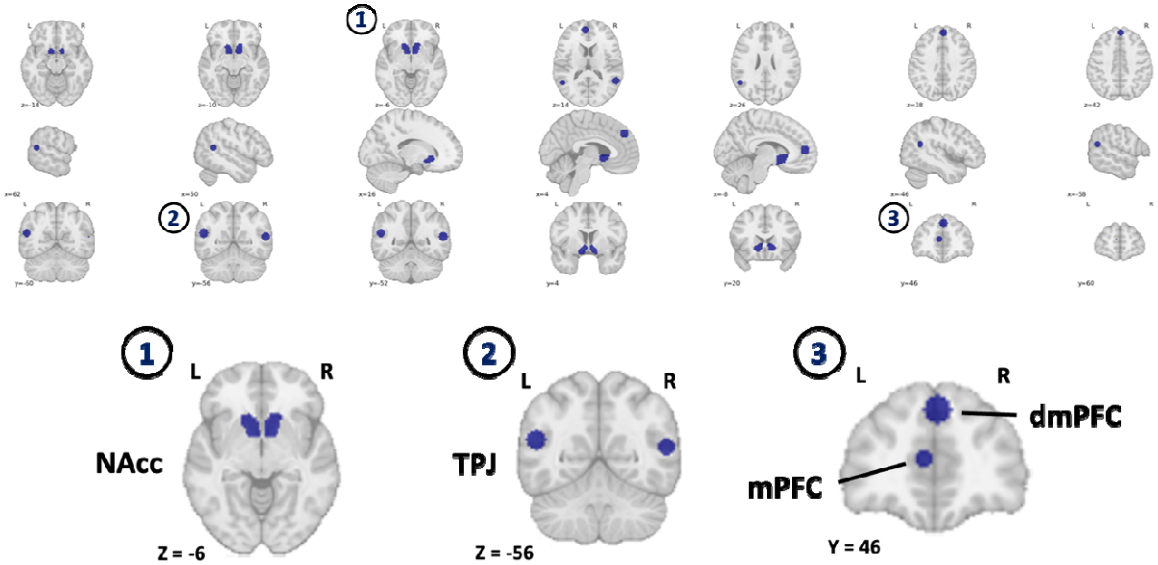
Condition x Self-parent neural dissimilarity x Linguistic preference score	-0.24 [-0.82, 0.35]	-0.11 [-0.77, 0.53]	-0.24 [-0.79, 0.32]	*-0.50 [-1.10, 0.08]*
Condition x Self-friend neural dissimilarity x Linguistic preference score	-0.64 [-1.07, 0.20]	0.45 [-0.26, 1.12]	0.08 [-0.56, 0.73]	-0.21 [-0.85, 0.39]

Supplementary Table 5. Predicting social decision preferences with interactions between linguistic preference scores and neural dissimilarity scores (TPJ, NAcc)

Term	TPJ – Monetary	NAcc – Monetary	TPJ – Social	NAcc – Social
Intercept	-2.46 [-3.00, -1.95]	-2.69 [-3.31, -2.07]	-0.85 [-1.27, -0.42]	-0.99 [-1.44, -0.54]
Condition	-0.03 [-0.62, 0.51]	0.13 [-0.47, 0.73]	-0.38 [-0.86, 0.17]	-0.27 [-0.82, 0.28]
Reward ratio	-2.91 [-3.52, -2.37]	-3.01 [-3.62, -2.42]	-1.29 [-1.62, -0.96]	-1.32 [-1.64, -0.98]
Self-parent neural dissimilarity (L)	0.14 [-0.31, 0.56]	0.06 [-0.48, 0.58]	-0.33 [-0.71, 0.04]	-0.11 [-0.47, 0.32]
Self-friend neural dissimilarity (L)	-0.63 [-1.08, -0.16]	0.10 [-0.49, 0.65]	-0.52 [-0.91, -0.13]	-0.37 [-0.82, 0.05]
Self-parent neural dissimilarity (R)	0.63 [0.17, 1.08]	0.00 [-0.54, 0.53]	0.17 [-0.22, 0.56]	0.27 [-0.14, 0.67]
Self-friend neural dissimilarity (R)	0.74 [0.29, 1.18]	-0.06 [-0.56, 0.46]	0.49 [0.11, 0.86]	-0.02 [-0.39, 0.36]
Linguistic preference score	0.10 [-0.35, 0.50]	0.15 [-0.38, 0.68]	-0.11 [-0.45, 0.23]	0.21 [-0.19, 0.63]
Condition x Self-parent neural dissimilarity (L)	-0.25 [-0.84, 0.35]	-0.30 [-0.92, 0.30]	-0.12 [-0.70, 0.45]	0.06 [-0.49, 0.66]
Condition x Self-friend neural dissimilarity (L)	-0.00 [-0.64, 0.66]	-0.26 [-0.93, 0.41]	-0.15 [-0.77, 0.45]	-0.11 [-0.76, 0.50]
Condition x Self-parent neural dissimilarity (R)	-0.32 [-0.94, 0.31]	-0.06 [-0.71, 0.53]	0.37 [-0.21, 1.00]	-0.54 [-1.15, 0.04]
Condition x Self-friend neural dissimilarity (R)	-0.35 [-0.96, 0.27]	0.70 [0.10, 1.33]	0.02 [-0.54, 0.61]	0.07 [-0.49, 0.66]
Condition x Linguistic preference score	-0.57 [-1.16, 0.00]	-0.47 [-1.08, 0.16]	0.30 [-0.23, 0.84]	-0.01 [-0.60, 0.60]
Self-parent neural dissimilarity (L) x Linguistic preference score	0.36 [-0.12, 0.87]	0.08 [-0.55, 0.68]	0.25 [-0.19, 0.66]	0.28 [-0.18, 0.75]
Self-friend neural dissimilarity (L) x Linguistic preference score	-0.47 [-1.02, 0.12]	0.51 [-0.04, 1.09]	-0.40 [-0.87, 0.07]	0.31 [-0.07, 0.71]
Self-parent neural	-0.12 [-0.60, 0.40]	-0.02 [-0.52, 0.51]	-0.11 [-0.54, 0.31]	0.07 [-0.33, 0.48]

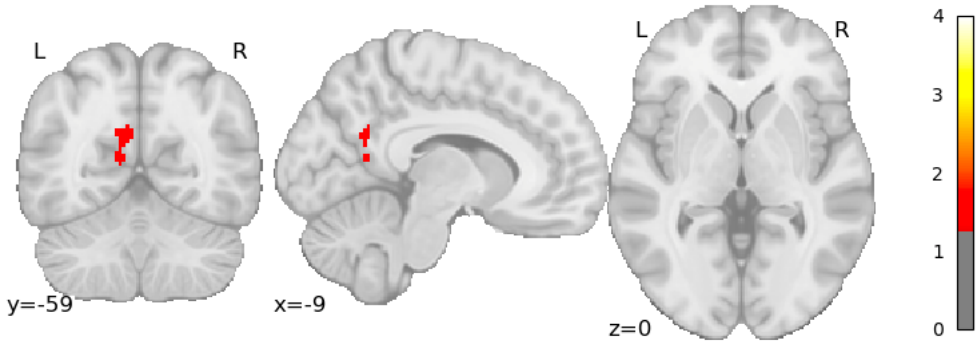
dissimilarity (R) x Linguistic preference score				
Self-friend neural dissimilarity (R) x Linguistic preference score	0.34 [-0.18, 0.94]	0.21 [-0.34, 0.76]	0.15 [-0.32, 0.63]	0.20 [-0.23, 0.61]
Condition x Self-parent neural dissimilarity (L) x Linguistic preference score	0.06 [-1.00, 0.35]	-0.16 [-0.88, 0.53]	-0.04 [-0.67, 0.61]	0.15 [-0.54, 0.84]
Condition x Self-friend neural dissimilarity (L) x Linguistic preference score	-0.02 [-1.29, 0.24]	*-0.62 [-1.27, 0.00]*	0.53 [-0.20, 1.24]	-0.35 [-0.92, 0.23]
Condition x Self-parent neural dissimilarity (R) x Linguistic preference score	-0.34 [-0.63, 0.72]	0.25 [-0.34, 0.88]	-0.14 [-0.77, 0.53]	*-0.48 [-1.08, 0.09]*
Condition x Self-friend neural dissimilarity (R) x Linguistic preference score	-0.53 [-0.79, 0.73]	-0.33 [-0.98, 0.33]	0.03 [-0.68, 0.76]	-0.16 [-0.79, 0.46]

Supplementary Figure 1. The six masks used for ROI analyses.



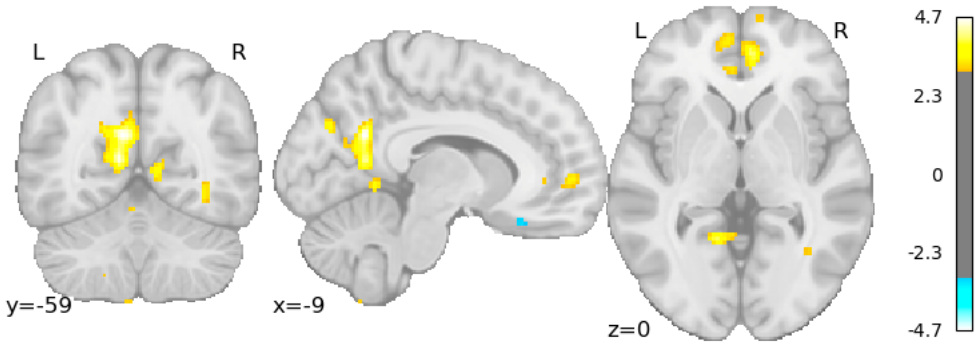
Note. Masks for the mPFC, dmPFC, and bilateral TPJ were created using Neurosynth. Bilateral NAcc masks were created using the Harvard-Oxford probabilistic atlas.

Supplementary Figure 2. Exploratory searchlight results showing distinct representations of parents and friends



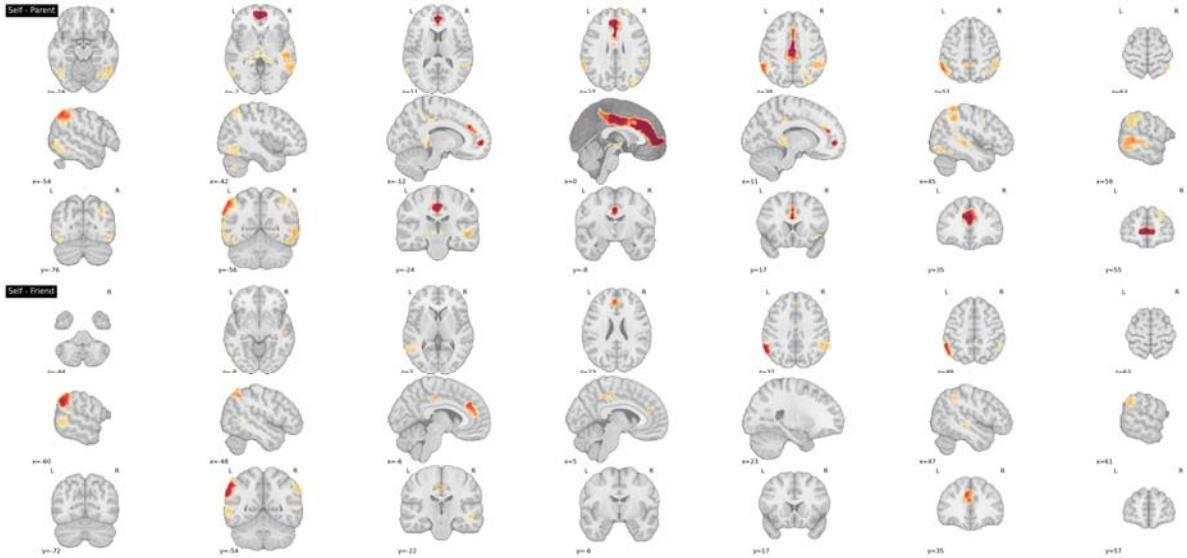
Note. Group level analyses were conducted via permutation test (10,000 iterations) using Nilearn's non_parametric_inference function and TFCE (FWE, $p < .05$). Negative log10 p-values are displayed as cluster intensity values. *P*-values are corrected for multiple comparisons (FWE, $p < .05$). The color bar ranges correspond to the threshold needed to achieve FWE error control (1.3) and the lowest possible *p*-value given the number of permutations (4).

Supplementary Figure 3. Uncorrected results of the exploratory searchlight analysis probing for distinct representations of parent and friend representations.



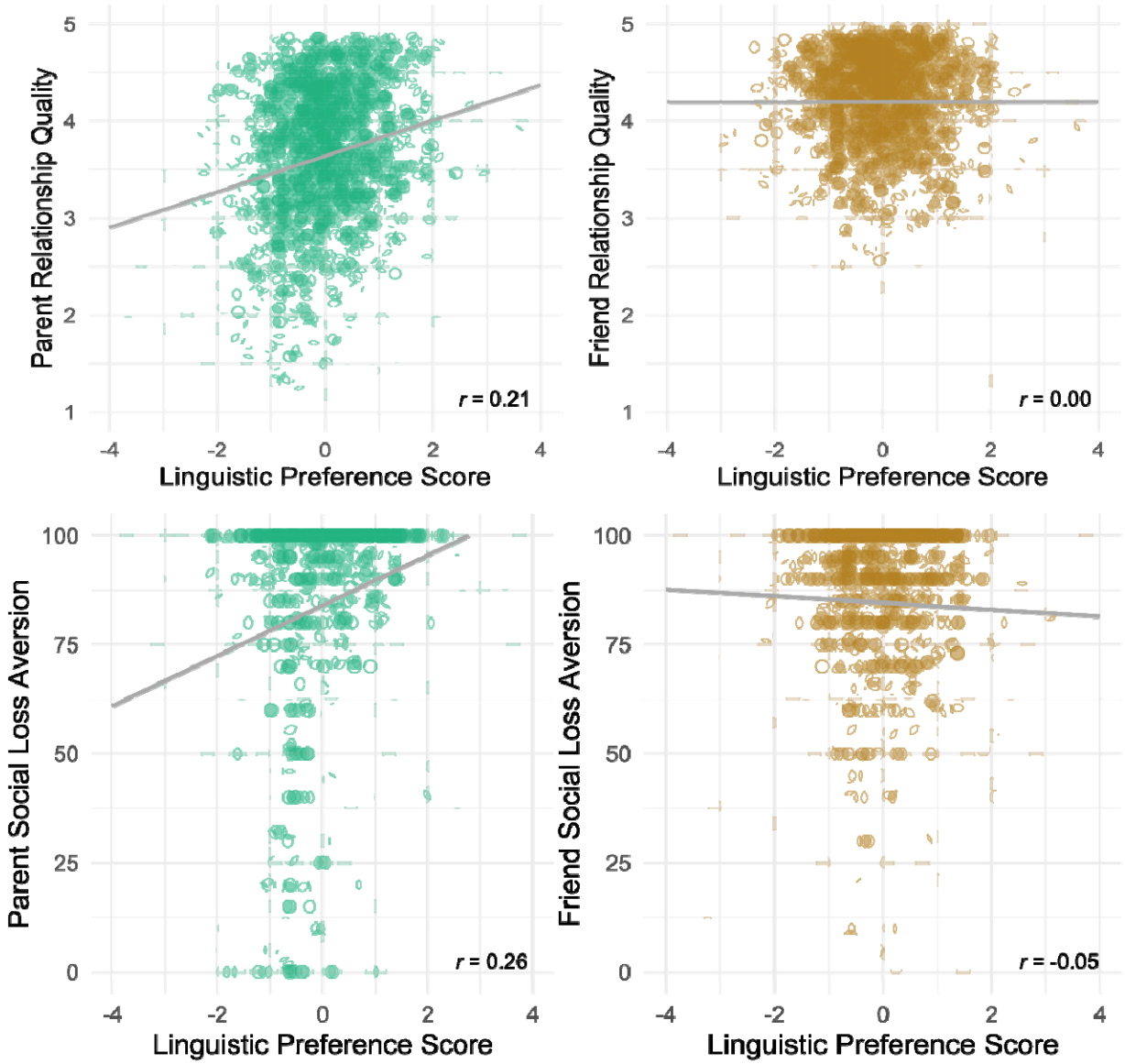
Note. Group level analyses were conducted via permutation test (10,000 iterations) using Nilearn's non_parametric_inference function. *t*-statistics are displayed for cluster intensity values (thresholded at 3.1). No corrections for multiple comparisons have been made to this image, it is plotted purely for exploratory purposes to see which clusters may have missed significance.

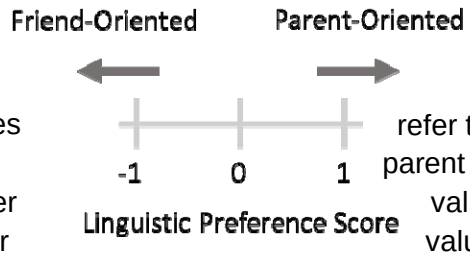
Supplementary Figure 4. Univariate contrasts for the social preference and judgment task.



Note. Group level analyses were conducted via permutation test (10,000 iterations) using Nilearn's non_parametric_inference function and TFCE. Negative log10 p-values are displayed as cluster intensity values. Clusters depict FWE-corrected *p*-values ($p < .05$).

Supplementary Figure 5. Linguistic preference scores strongly generalize to other relationship characteristics (relationship quality, social loss aversion) for parents, weakly generalize to friend social loss aversion.





Note. Linguistic preference scores refer to model implied social decision preferences between a parent and friend based on written text data; relatively greater values indicate a parent-over-friend preference, relatively lower values indicate a friend-over-parent preference. The Y-axis depicts relationship quality between participants and their nominated parent (left) and friend (right). Relationship quality was obtained with the self-administered IPPA. Social loss aversion is a one-shot item where participants are asked how

upset they would be if they could no longer spend time with a given other. Linguistic preference scores were calculated with weights trained on the Word2Vec-based linguistic representations.