Explaining Individual Differences in Advantageous Inequity Aversion by Social-Affective Trait Dimensions and Family Environment

Hongbo Yu[1] , Chunlei Lu[2], Xiaoxue Gao[3], Bo Shen[2], Kui Liu[2],
Weijian Li

dissociable underlying neurocognitive mechanisms (Gao et al., 2018; R. Yu et al., 2014). One way to further probe the underlying mechanisms is to examine the social and affective factors associated with advantageous and disadvantageous IA.

Although prior research on fairness-related behaviors and IA have documented individual differences of advantageous and disadvantageous IA (Engel, 2011; Gao et al., 2018; Tisserand et al., 2015; R. Yu et al., 2014), systematic investigations into the core social and affective factors that may distinguish advantageous and disadvantageous IA have been lacking. One possibility is that these two types of IA rely differently on the ability and tendency to take others' well-being into consideration (Tsoi & McAuliffe, 2020). Specifically, we hypothesized that other-regarding social-affective traits (e.g., empathic concern and guilt) and their antithesis (e.g., callousness and interpersonal manipulation) are associated with advantageous IA but not disadvantageous IA. We note that advantageous and disadvantageous IA in the strictest sense is a description of certain behavioral patterns in a specific economic game, and it may or may not be associated with aversive emotional responses (Binmore & Shaked, 2010; Fehr & Schmidt, 2010). We use these terms in their descriptive sense.

Other-regarding social-affective traits, such as empathic concern and guilt proneness, predispose individuals to be more attentive to the distress of others and motivate altruistic behaviors (Blair & Mitchell, 2009; Kimonis et al., 2019; Thielmann et al., 2020). For example, past research has shown that individuals with high guilt proneness are less likely to commit unethical behaviors (e.g., lying in negotiation) and are more likely to take reparative measures after transgression (Cohen et al., 2012; Cohen et al., 2011; Giner-Sorolla et al., 2011; Tangney et al., 2000). Similarly, numerous empirical studies have offered supportive evidence for the empathy-altruism hypothesis, which posits that at least some forms of empathy motivate observers to help victims for the sake of the victims' well-being (Batson, 2011; Batson et al., 2007; Davis, 2015; Dovidio et al., 1990; Penner et al., 2005; Stocks et al., 2009; Wilhelm & Bekkers, 2010; Zaki, 2019, 2020).

In contrast, deficits in prosocial affective traits, both in clinical and general populations, have been associated with antisocial behaviors and a lack of care and altruistic responses to others' distress (Blair, 2008, 2013; Blair et al., 2005; Glenn & Raine, 2014; Gregory et al., 2015; Yang et al., 2015). For instance, in general adult populations, self-reported psychopathic features have been associated with higher tendency to gain financial reward for oneself by harming another person in laboratory settings (Crockett et al., 2014), as well as more violent/aggressive behaviors in everyday life (Neumann & Hare, 2008). In adolescents, callous-unemotional (CU) traits have been shown to predict conduct problems, criminal offending, and delinquency (Frick & Viding, 2009). A recent meta-analysis reveals that CU traits are strongly and negatively associated with prosocial behavioral traits (Waller et al., 2020).

Although past research has established the link between social-affective traits and prosocial behaviors (or the lack thereof), two questions remain unclear. First, these previous studies have been primarily focused on behavioral outcomes and therefore remain agnostic about what underlying cognitive processes are influenced by prosocial emotions and social-affective traits. A seemingly prosocial behavior (e.g., allocating resources fairly) may be driven by multiple, sometimes conflicting underlying cognitive processes (e.g., an aversion toward inequality, guilt when getting more than one should, envy when the other party receives more than oneself). Specifically, in one of the most widely adopted experimental tasks for probing prosociality, namely, the dictator game (DG; Engel, 2011; Forsythe et al., 1994), participants decide how they want a pool of money to be divided between themselves and a receiver, while the receiver has to accept what is allocated to them. In a modified version of DG, participants face a series of binary choice in which one option is always a fixed fair division, while in the other option, the amount for the participants and the amount for the receiver are orthogonalized (for details, see Methods and Materials section). Combined with an established computational model for IA, this paradigm has the advantage of statistically dissociating advantageous and disadvantageous IA, thereby allowing us to examine the factors that drive individual differences in these two latent cognitive processes (Fehr & Schmidt, 1999; Gao et al., 2018).

Second, past research on the individual differences in fairness-related behaviors with adult populations has largely overlooked the link between the social-affective traits predictive of prosocial behaviors and features of the environment where these traits develop. The environments where individuals socialize (e.g., the individuals' family environment) play a key role in the development of the individuals' personality (Di Pierro et al., 2012; Hoffman, 1991; Loehlin & Nichols, 2012). An understanding of potential environmental antecedents of these social-affective traits may have implications for interventions aimed to bolster the development of the traits that are conducive to prosocial behaviors and discourage those that may hinder prosocial behaviors (Singer & Klimecki, 2014). To fill this gap, we explored the effect of a potential environmental factor—the extent to which family members are encouraged to express their thoughts and feelings, support, care, and empathize with each other (Ferguson & Stegge, 1995; Hinde, 2002; Stuewig & McCloskey, 2005; Tangney & Dearing, 2003). Indirect evidence from research on the relationships between parenting style and children's and adolescents' empathy-related traits, guilt proneness, and prosociality suggests that positive and emotionally responsive parenting facilitates the development of empathy, care, and a sense of guilt (Eisenberg & Valiente, 2002; Kochanska, 1991, 1997; López et al., 2008; Miklikowska et al., 2011). In this study, we test the prediction that a supportive family environment, as measured by a self-reported questionnaire (see Methods and Materials section for detail; Kang et al., 2001), is associated with more other-regarding social-affective traits, which in turn results in higher advantageous, but not disadvantageous, IA.

We carried out two studies to better understand the social-affective factors that drive individual differences in

advantageous (relative to disadvantageous) IA. In Stud 1, e administered the modified DG to a sample of incarcerated adolescents (N = 67). The rationale of including this sample as to ma imi e the range of the distribution of CU traits, as it has been demonstrated that these traits have a ider distribution in institutionali ed samples than in the general population (B rd et al., 2013; Essau et al., 2006; Kimonis et al., 2008; Pihet et al., 2015). In Stud 2, e aimed to replicate and e tend the findings from Stud 1 in a large sample of undergraduate students (N = 2,250) in a Chinese universit . The si e of this sample allo s us to adopt a dimensional (or trans-diagnostic) approach to personalit traits in computational ps chiatr (Gillan et al., 2016), running factor anal sis on individual items from various partiall overlapping questionnaires and using the resultant factor scores, rather than questionnaire total scores, as predictors of the latent cognitive processes underl ing the DG choices.

Methods and Materials

Participants

Study 1. To e amine ho CU trait modulates prosocial motivation, e paid a visit to a correctional institution and administered a resource allocation task (modified DG; see belo) to a group of 67 incarcerated male adolescent participants (mean age: 16.3 ± 0.8, age range = 14 17) in the institution. In the countr here the data ere collected, a correctional institution is a t pe of confined facilit for juvenile offenders under the age of 18. The t pes of criminal offenses this sample of participants committed can be found in Online Supplemental Table S1. The procedure used in the present stud as approved b the authors' universit ethics committee and as administered as part of the institution's ps chological intervention program. The si e of the incarcerated sample as determined b ho took part in the institution's ps chological intervention program and as available on the da of data collection.

Study 2. First- ear undergraduate students at a universit in southeast China participated in the stud as part of the universit 's mental health prescreening. The stud as approved b the authors' universit ethics committee. Participants gave their consent electronicall prior to the e periment. A total of 4,888 participants completed the stud as part of their mental health assessment program mandated b their universit authorit . Among them, 2,638 participants ere e cluded from data anal sis due to failure in comprehension or attention check questions, leaving a sample of N = 2,250 (mean age: 18.2 ± 0.7; age range = 17 22; 1,679 ere female; see belo for detailed e clusion criteria). Note that adopting different e clusion criteria does not change the pattern of results (see Online Supplementar Material, p. 8). The sample si e as determined b the number of first- ear undergraduate students at the universit here the data collection took place.

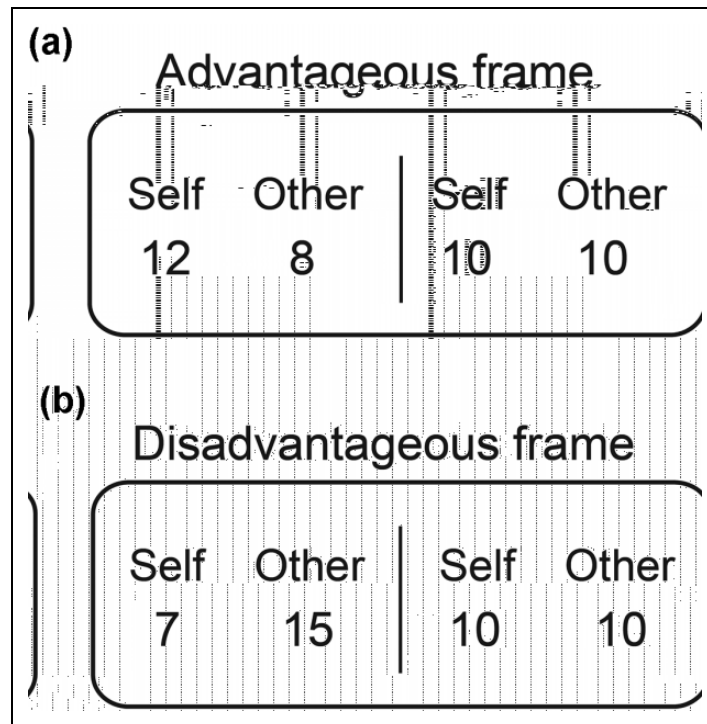**Table 1.** Demographic and Personality Measures of the High and Low Callous-Unemotional (CU) Groups.

Measure	High CU M (SD)	Low CU M (SD)	t Value	p Value
Age	16.3 (0.8)	16.3 (0.8)	−0.18	.86
Education	2.1 (1.1)	2.0 (0.5)	0.30	.77
Father education	1.8 (1.0)	1.9 (1.3)	−0.56	.58
Mother education	2.0 (1.5)	1.6 (0.7)	1.24	.22
Family gross income	3.3 (1.5)	3.7 (1.4)	−1.02	.31
Callous-unemotional	10.3 (1.5)	8.3 (2.1)	4.63	<.001
IRI-perspective taking	1.9 (0.6)	2.2 (0.5)	−2.08	.04
IRI-empathic concern	2.1 (0.6)	2.6 (0.4)	−3.98	<.001
IRI-personal distress	2.2 (0.5)	2.1 (0.5)	0.98	.33

Note. IRI = Interpersonal Reactivity Index.

Experimental Design and Measurements

Overview. For Stud 1, participants' CU traits ere assessed prior to this e periment session via intervie s conducted b trained research assistants (Essau et al., 2006; Kimonis et al., 2008; the Chinese version of the assessment scale as adopted from Chen [2013]). A high CU group (N = 32) and a lo CU group (N = 35) ere defined based on median split of the overall CU score (Table 1; cf. Pihet et al., 2015). A computer program as installed in the computers in the testing room. This program ould present the DG task to the participants and record their responses (i.e., button press). For Stud 2, participants first performed the modified DG task ith an anon mous co-pla er ho as also a participant in the same stud session. The participants then completed several personalit questionnaires and provided demographic information (see belo for details). The task and the questionnaires ere computeri ed and presented to the participants via an online surve platform (https:// . j .cn/).

Modified DG. We instructed the participants that the ould be paired ith an anon mous co-pla er in the same room. The participants' task as to allocate monetar points bet een themselves (hereafter, self) and the anon mous co-pla er (hereafter, other), in the form of binar choice (Figure 1; for participant pa ment, please see Online Supplementar Material, p. 6). Specificall , one of the t o options al a s offered 10 points to each pla er. The other option came from a set of test options var ing in the pa off of self (**Ms**) and the pa off of other (**Mo**; see Online Supplemental Figure S1 for the full list of options used in this stud). Participants made a series of 50 choices and one of them ould be randoml selected and made real at the end of the e periment. From the participants' perspective, in 48 of the 50 trials, the test option as either advantageous (i.e., **Ms** > **Mo**; Figure 1a) or disadvantageous (i.e., **Ms** < **Mo**; Figure 1b). The test options ere generated, such that **Ms**, **Mo**, and the absolute inequit (i.e., **Ms** **Mo**) ere decorrelated (rs < .07, ps > .66; Gao et al., 2018; Sae et al., 2015). T o of the 50 trials ere catch trials here

Figure 1. Bi a ch ice i he dified dic a ga e. O e f he i a al a a fi ed, fai di i i he e b h elf a d he ld ge 10 i . The he i ca e f a e f e i a i gi he a ff f he a ici a he el e (*Ms*) a d he a ff f he (*Mo*). E a le ial f he ad a age f a e (a) a d he di ad a age f a e (b) a e h .

the test option was also a fair division. Specifically, $Ms = Mo = 2$ for one, and $Ms = Mo = 18$ for the other.

Computational modeling of choice in the modified DG. We modeled participants' trial-by-trial choices by adapting a two-player IA model (Fehr & Schmidt, 1999; see also Charness & Rabin, 2002) that had been validated for these types of binary choice task (Gao et al., 2018; Sáez et al., 2015). This allowed us to quantitatively isolate two motivations underlying participants' choices:

$$U = Ms - q \cdot \alpha \cdot (Mo - Ms) - p \cdot \beta \cdot (Ms - Mo),$$

where *Ms* and *Mo* are participants' payoff and the recipient's payoff in a given option, respectively. *p* and *q* indicate whether the option involves advantageous inequity or "aheadness" ($p = 1$ when $Ms > Mo$, $p = 0$ otherwise) or disadvantageous inequity or "behindness" ($q = 1$ when $Mo > Ms$, $q = 0$ otherwise). Note that in some literature, the meaning of *p* and *q* is reversed (Gao et al., 2018). α and β are free parameters indicating the degree of disadvantageous IA and advantageous IA, respectively. We used a softmax function to convert utility difference between the two options ($\Delta U = U_{\text{unequal}} - U_{\text{equal}}$) into probability of choosing the unequal option:

$$P(\text{unequal}) = \frac{1}{1 + e^{-\lambda \Delta U}}.$$

Here, the inverse temperature parameter (λ) captures the steepness of the softmax function: Higher value means that the

softmax curve is closer to a step function, which in turn indicates that the participant's choice is more sensitive to the change in utility difference. Given the relatively small sample and noisier choice behaviors of the incarcerated sample (see Online Supplementary Material, p. 7), we only estimated α and β at the group level for Study 1 (cf., Gao et al., 2018; Zhu et al., 2014). For Study 2, we estimated α and β for each individual participant. A maximal likelihood estimation was used to find the combination of free parameters that best fit the observations. Two hundred iterations were performed for the group-level estimation (Study 1), while 50 iterations were performed for each individual participant (Study 2). The computational model quite accurately predicted the participants' choices (69% for the high CU participants, 68% for the low CU participants, and 85% for the college student sample).

Self-reported personality questionnaires and demographic information. For Study 1, participants completed the Interpersonal Reactivity Index (IRI; Davis, 1983) on a separate day prior to the experimental session. For Study 2, participants completed a battery of personality questionnaires assessing their social-affective traits, including the IRI, a 30-item Self-Reported Psychopathy Scale (Bartels & Pizarro, 2011), the Guilt and Shame Proneness Scale (GASP; Cohen et al., 2011; Young et al., 2019), the Self-Compassion Scale (Neff, 2003), and the Toronto Alexithymia Scale (Bagby et al., 1994). Participants' attitudes and beliefs regarding justice and fairness were assessed using the General Belief in a Just World Scale
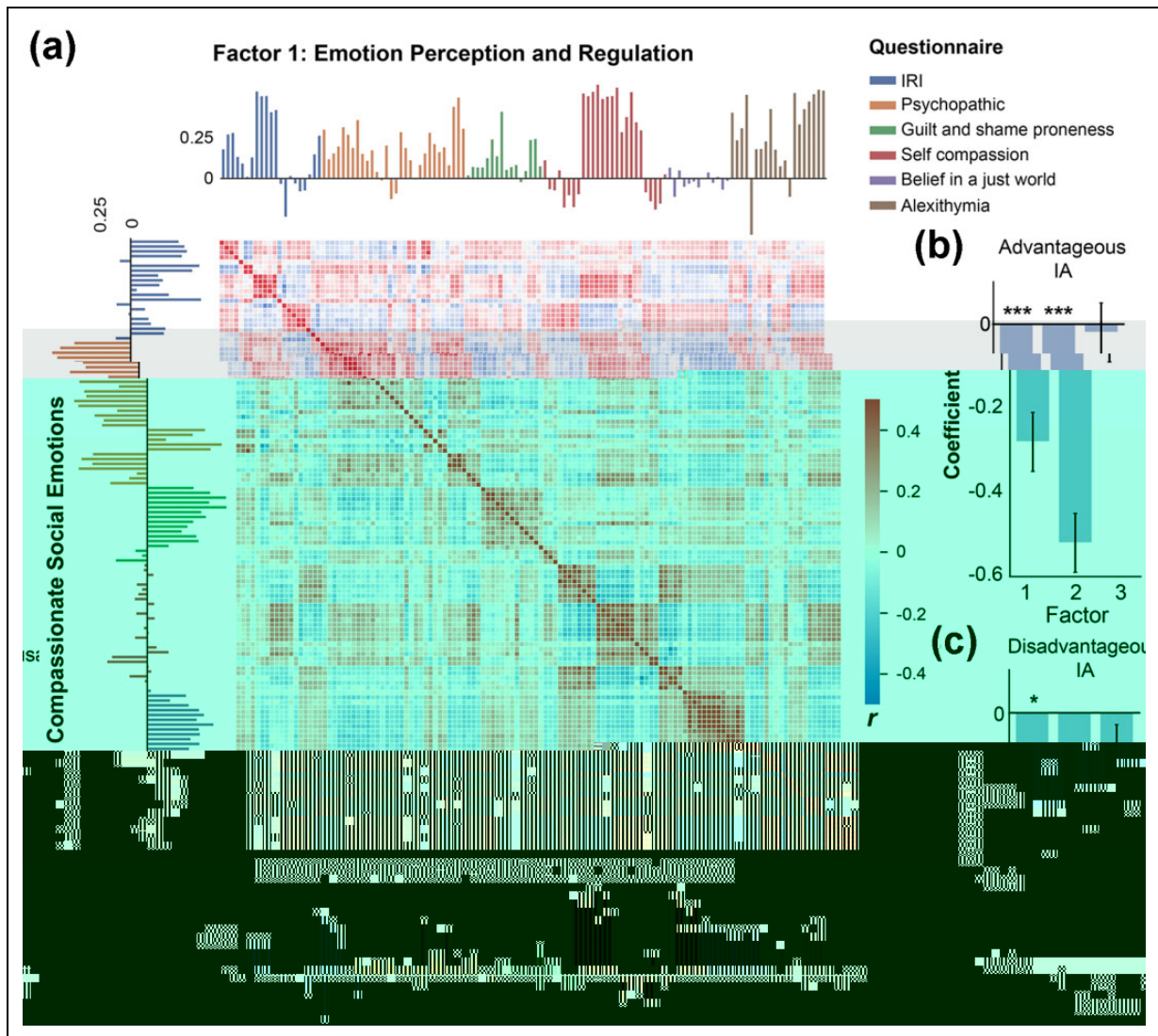
Figure 2. Results of factor analysis. (a) The correlation matrix of 126 individual questionnaire items and loadings of each item for the three factors. (b and c) Factor scores of each participant were entered into linear regression models for advantageous and disadvantageous inequity aversion (IA) parameters. Advantageous IA was negatively associated Factors 1 and 2. Disadvantageous IA was only significantly associated with Factor 1. Error bars indicate s.e.m. *$p < .05$. ***$p < .001$. s.e.m. = standard error of the mean.

(Da be , 1999). T e Se f-Ra   Sca e f S   e c Fa
D   c (SSFD; Ka   e a ., 2001) a   d ded   a e
 a c a ' e ce   f e fa e   e .T e
SSFD c a ac e e e   a a   a d a e   f c   -
ca   a d e ac   a fa   .T e ca e a f
d e   (ee O   e   e e a Ma e a , .2, f   e
    a C e e   a d E   a a) a d e
 ec f ca   f   ed   e Fa   A   e e (FA)   b ca e,
  c d ca e a ca   a d   e FA (e ., M fa
 e be ca e a e   e   a   a d c   e f eac
 e). Pa c a a   ded de   a   c f a ,
   d   e a e, e a   e a b ,   e   e a e
a   c d   e fa   , e e ed ca   f e a-
e , e e   e   e e e   (ba   , a), a d
 e   b ec e   ca e c   c a .T e e a a be e e
 ed a c   a a e   da a a a e .

Data exclusion criteria for Study 2. I   a , e   d ded   ee
 ec a   a e   e a   e a c a de a d e
DG a   a d a a   ff c e a e   ,   e a
(O   e S   e e a Ma e a , . .5). F , af e e   a c-
 a   ead e   c   f   e DG a , e eed a e
e   c   e e   e   ab   e DG a . Sec d, e
 d ded   ca c a   e DG a , e e e
b   e   f a b e a   e   e b   f   e   a c-
 a (.e., dec de) a d e ec   e .T d, e e ed ee
 a e   c ec   e   e e   a   e   a e a
 e e b   a d   b ec e. Pa c a   c   ec
 a   e ed a   ec   e e   c ec   e   e DG a
 a d ea e   c ec   e e a   e a e ee
  d ded   da a a a   . .0   e   a e a   de ca   de
d ffe e   da a a c   c e a (O   e S   e e a
Ma e a , . .8).

Table 2. [text corrupted/unreadable]

Variable			B (SE) CI Advantageous IA	Disadvantageous IA B (SE) CI	Inverse Temperature B (SE) CI
Factor 1: E			−.28 (.07)*** [−0.42, −0.13]	−0.23 (.10)* [−0.42, −0.04]	.03 (.01)* [0.01, 0.05]
Factor 2: C			−.52 (.07)*** [−0.66, −0.38]	−0.16 (.10) [−0.35, 0.30]	.06 (.01)*** [0.03, 0.08]
Factor 3: E			−.02 (.07) [−0.17, 0.12]	−0.12 (.10) [−0.32, 0.07]	.01 (.01) [−0.01, 0.04]
Sex (>)			−.52 (.16)** [−0.84, −0.20]	−1.44 (.22)*** [−1.87, −1.007]	.19 (.03)*** [0.14, 0.24]
A			.12 (.11) [−0.09, 0.33]	0.07 (.15) [−0.22, 0.35]	−.03 (.02) [−0.06, 0.01]
S (>)			−.15 (.15) [−0.44, 0.13]	−0.53 (.20)** [−0.91, −0.15]	.05 (.02)* [0.01, 0.09]
U (>)			.11 (.17) [−0.21, 0.44]	0.07 (.22) [−0.37, 0.51]	−.01 (.03) [−0.06, 0.04]
F			−.08 (.07) [−0.22, 0.06]	−0.21 (.10)* [−0.40, −0.03]	.02 (.01) [−0.06, 0.04]
M			−.08 (.07) [−0.22, 0.06]	0.00 (.10) [−0.18, 0.19]	−.01 (.01) [−0.03, 0.01]
S			−.05 (.05) [−0.15, 0.05]	0.12 (.07) [−0.02, 0.26]	−.00 (.01) [−0.02, 0.01]
N			2,250	2,250	2,250
R^2			.05	.05	.07

[footnote text corrupted] ... CI =
*p < .05. **p < .01. ***p < .001.

Results

All deidentified data and data analysis codes related to the results reported in this article can be accessed at https://osf.io/fge9v. We have reported all measures, conditions, data exclusions, and how we determined the sample sizes.

In Study 1, we tested the hypothesis that high CU trait is associated with advantageous, but not disadvantageous, IA. Supporting this hypothesis, the advantageous IA of the high CU group ($M \pm SD = 0.75 \pm 0.07$, credible interval [89% highest density interval] = [0.662, 0.863]) was almost 50% lower than that of the low CU group (1.14 ± 0.09, credible interval = [1.007, 1.253]; Online Supplemental Figure S2a and Table S2). This was not the case for disadvantageous IA (high CU group: 0.52 ± 0.10, credible interval = [0.344, 0.655]; low CU group: 0.59 ± 0.12, credible interval = [0.420, 0.773]; Online Supplemental Figure S2b). This pattern indicates that getting more than one's fair share is less of a concern for individuals with high CU than those with low CU, but they are equally averse to getting less than their fair share. The inverse temperature parameter of the high CU group (0.13 ± 0.02, credible interval = [0.116, 0.168]) was higher than that of the low CU group (0.10 ± 0.01, credible interval = [0.079, 0.115]; Online Supplemental Figure S2c).

In Study 2, we aimed to (1) conceptually replicate the differential effects of callousness-related traits on advantageous versus disadvantageous IA and (2) to examine the specificity of the effects of callousness-related traits in a larger noninstitutionalized sample. We found that the scores of the Callous Affect and Interpersonal Manipulation subscales of the self-reported psychopathy questionnaire (Bartels & Pizarro, 2011) were strongly and negatively correlated with advantageous IA. Moreover, the correlations with advantageous IA were significantly stronger (i.e., more negative) than those with disadvantageous IA (Online Supplemental Table S3). This pattern, however, was not specific to callousness-related traits. In fact, most of the social-affective personality traits that we measured showed a similar pattern (for details, see Methods and Materials section and Online Supplemental Table S3). Given the conceptual and statistical overlap among the questionnaires, including their total scores in the same regression model to predict IA parameters is both uninformative and problematic.

To address this issue, we adopted a dimension approach to personality measures (Gillan et al., 2016) and used the composite dimensional scores to predict participants' behavioral preferences in the DG task. Specifically, we carried out a factor analysis on the 126 individual items from the six personality questionnaires. Using the Cattell–Nelson–Gorsuch test implemented by the "nFactors" package in R (Raiche & Magis, 2010), our analysis identified a three-factor latent structure (Figure 2a). Based on the highest loading items (|loading| > 0.25), we labeled the factors as "emotion perception and regulation" (Factor 1; Online Supplemental Table S4, an example item "Being in a tense emotional situation scares me,"

loading = 0.54), "compassionate social emotions" (Factor 2; Online Supplemental Table S5, an example item "I often have tender, concerned feelings for people less fortunate than me," loading = −0.31), and "expanded self and belief in justice" (Factor 3; Online Supplemental Table S6, example items "I try to see my failings as part of the human condition," loading = 0.57 and "I think basically the world is a just place," loading = 0.32).[1] Of particular interest, "compassionate social emotions" (Factor 2) picked up almost all the individual items from the Interpersonal Manipulation ($M \pm SD$ = 0.43 ± 0.06) and the Callous Affect (0.34 ± 0.11) subscales of the self-reported psychopathic questionnaire, and all the items from the GASP scale pertaining to guilt (−0.40 ± 0.09) and shame experience (−0.39 ± 0.09; Online Supplemental Table S7). Therefore, higher scores on this dimension indicate a **lack** of dispositional compassionate social emotions.

We next ran two robust linear mixed-effect models (R package "robustlmm"; Koller, 2016) to examine the association between the factor scores and the advantageous and disadvantageous IA parameters, which were estimated independently of the factor analysis. The scores of all the three factors were included in the same model. Demographic variables were also included as covariates (Table 2). Both Factor 1 and Factor 2, but not Factor 3, were significantly and **negatively** associated with the advantageous IA parameter (Table 2, Online Supplemental Table S8, and Figure 2b). For the disadvantageous IA parameter, only Factor 1 was significantly correlated (Table 2, Online Supplemental Table S8, and Figure 2c). Note that this latter association became nonsignificant under the most conservative data exclusion criteria, indicating that this effect was not as robust as the effects with the advantageous IA. Importantly, as the confidence intervals indicated, Factor 2 was significantly more predictive of advantageous IA than of disadvantageous IA. This differential predictive power, which was conceptually consistent with the finding of Study 1, was not observed for Factor 1 or Factor 3. We carried out a post hoc power calculation based on the association between Factor 2 score and advantageous IA (f^2 = 0.014). The size of the final analysis sample afforded a power of 99.9% in detecting this effect at **p** < .05.

Finally, we explored whether a supportive family environment leads to higher advantageous IA, via the mediating role of the social-affective factors that were predictive of advantageous IA. Family environment was indicated by the scores on the FA subscale of the Self-Reported Family Dynamics Scale (Kang et al., 2001). This subscale reflects the degree to which one's family is caring and supportive to its members (Cronbach's α = .89). As Online Supplemental Table S3 shows, FA score was significantly positively correlated with advantageous IA and significantly more so than with disadvantageous IA. We ran a mediation model where FA score was entered as the independent variable, the scores of Factor 1 and Factor 2 as two parallel mediators, and the advantageous IA parameter as the dependent variable. An SPSS macro was used to evaluate mediation models (Hayes, 2013). For the mediation analysis, we also included covariates of no interests as stated

above (Table 2). We found that Factor 1 and Factor 2 together fully mediated the relationship between FA and advantageous IA (direct effect: **B** = −.19, **SE** = .13, CI [−0.45, 0.07]). The mediation effects of Factor 1 (**B** = .12, **SE** = .03, CI [0.06, 0.19]) and Factor 2 (**B** = .34, **SE** = .05, CI [0.24, 0.45]) were significantly above zero (Figure 3). The mediation effect of Factor 2 was significantly stronger than that of Factor 1 (mean difference = 0.22, **SE** = .06, CI [0.09, 0.35]). Note that although both of the mediation effects are positive, they are "inhibitory," meaning that a positive family environment is negatively associated with the personality dimensions that are themselves "inhibitors" of advantageous IA, therefore "disinhibit" it.

Discussion

Utilizing a computational model, we demonstrated in a sample of incarcerated adolescents the contribution of CU traits to advantageous IA. In a follow-up study with a large-scale college student sample (**N** = 2,250), we conceptually replicated and extended the association between callousness and advantageous IA by adopting a dimensional approach to social-affective personality traits. We found that a trait dimension characterized as "compassionate social emotions" was most predictive of advantageous IA but was unrelated to disadvantageous IA.

Past research has documented that individuals with high callousness or low GASP are more likely to engage in unethical behaviors (Blair, 2013; Cohen et al., 2012; Waller et al., 2015). Replicating and extending those previous studies, here we revealed that it was the advantageous IA underlying the prosocial behavior that was modulated by this social-affective trait dimension. Interestingly, this social-affective trait dimension was not associated with either diminished or heightened disadvantageous IA. This suggests that this social-affective trait dimension, which is primarily concerned with one's own unethical behaviors, is dissociable from envy and reactive aggression, which is primarily concerned with unfairness and injustice inflicted on oneself (Costa & Babcock, 2008; Meehan et al., 2001; Walker & Jackson, 2017). It is the former that depends on the agent's sense of shared social commitments and has only been found in older children and adult human beings, but not in younger children or nonhuman primates (McAuliffe et al., 2015; Tomasello, 2020; Tsoi & McAuliffe, 2020; Ulber et al., 2017). With these correlational analyses, however, we do not intend to overinterpret our results as implying any direction of causality.

The items traditionally included in the subscales of the IRI (i.e., personal distress, empathic concern, and perspective taking) nicely mapped onto different latent factors, suggesting their dissociable roles in motivating prosocial behaviors. Specifically, personal distress strongly and consistently loaded positively on Factor 1, which is associated with lower advantageous IA. This is consistent with ample empirical evidence that personal distress is self-centered and promotes withdrawal; even when it motivates helping behaviors, the underlying

motivation is more to terminate one's own distress than to benefit the recipient (Batson, 2011; Batson et al., 1981). In contrast, empathic concern loaded negatively on Factor 2, which is associated with advantageous IA. Previous research has demonstrated that empathic concern, unlike personal distress, is other-regarding and has an approach tendency (Davis et al., 1999; FeldmanHall et al., 2015; Zaki, 2014).

It is interesting to compare the effect of the trait dimension represented by Factor 2 and episodic social emotions (e.g., guilt) on IA. For example, Gao et al. (2018) has demonstrated that when episodic guilt state was induced experimentally, individuals exhibited higher advantageous IA *and* lower disadvantageous IA. This is conceivable because retrospective guilt should not only discourage individuals from engaging in future transgression but also motivate individuals to make amend for existing transgression and damage (De Hooge, 2019; Kamau et al., 2013; H. Yu et al., 2014). In contrast, many of our social-affective trait measures are anticipatory in nature (Cohen et al., 2012). Our result lends support to a cognitive account of the prosocial function of social affective traits (i.e., compassionate social emotions), namely, individuals who anticipate more future social emotions (e.g., guilt, shame) find

the prospect of unjustly getting better off than others more aversive (see also Gong et al., 2019). Future studies are necessary to ascertain the neurobiological links between behavioral tendency (e.g., advantageous IA), episodic social emotions (e.g., guilt), and social-affective traits (e.g., guilt proneness).

Our finding that a positive family environment is associated with social affective traits pertaining to compassionate social emotions provides evidence for the developmental observations that family environment and parental warmth play a key role in the proper development of prosocial emotions such as empathy and guilt (Ferguson & Stegge, 1995; Hinde, 2002; Tangney & Dearing, 2003; Zahn-Waxler & Kochanska, 1990). The novel contribution of our findings is that we revealed possible routes from family environment to prosocial behavioral preference via social-affective traits. However, it should be noted that these results are correlational and should be interpreted with caution. For example, the mediation results cannot rule out the possibility that participants low in social-affective trait have inaccurate and self-motivated perceptions or memories of their family interactions (Klein & Epley, 2016; Tasimi & Johnson, 2015). Rigorous developmental experiments are needed to establish the causal relationship

between positive family environment and compassionate social emotions.

To conclude, by combining computational modeling and a dimensional approach to personality measures, this well-powered study offers a cognitive account of how compassionate social emotions as a social-affective trait promotes prosocial behaviors—individuals high on this dimension are more careful not to be unfairly better off than others (i.e., advantageous IA). Moreover, we highlight the association between a positive family environment and the development of the trait of compassionate social emotions and provide evidence for an intermediate role of affective trait in the relationship between family environment and advantageous IA. Together, the results of this study suggest that the trans-diagnostic approach is not only useful in discovering dimensional markers of behavioral anomaly in psychiatry but is also applicable to ascertaining the specificity of social-affective trait dimension in predicting prosociality.

Acknowledgments

Declaration of Conflicting Interests

Funding

ORCID iDs

Hongbo Yu 🔟 https://orcid.org/0000-0002-3384-7772
Xudong Zhao 🔟 https://orcid.org/0000-0003-1493-3517

Supplemental Material

The supplemental material is available in the online version of the article.

Note

1. Note that, for the first two factors, the scores indicate the opposite or the lack of the traits signified by their respective factor labels. For example, higher score on Factor 2 indicates a lack of compassionate social emotions. We decided not to label the factors this way to avoid wordy names.

References

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. , 38(1), 23–32.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. , 121(1), 154–161. https://doi.org/10.1016/j.cognition.2011.05.010

Batson, C. D. (2011). A . Oxford University Press.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? , 40(2), 290.

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., & Ortiz, B. G. (2007). An additional antecedent of empathic concern: Valuing the welfare of the person in need. , 93(1), 65.

Binmore, K., & Shaked, A. (2010). Experimental economics: Where next? B & , 73(1), 87–100.

Blair, R. J. R. (2008). The amygdala and ventromedial prefrontal cortex: Functional contributions and dysfunction in psychopathy. B: B , 363(1503), 2557–2565.

Blair, R. J. R. (2013). The neurobiology of psychopathic traits in youths. , 14(11), 786.

Blair, R. J. R., & Mitchell, D. G. V. (2009). Psychopathy, attention and emotion. , 39(4), 543.

Blair, R. J. R., Mitchell, D., & Blair, K. (2005). : . Blackwell Publishing.

Brosnan, S. F., & de Waal, F. B. M. (2014). Evolution of responses to (un) fairness. , 346(6207), 1251776.

Byrd, A. L., Kahn, R. E., & Pardini, D. A. (2013). A validation of the Inventory of Callous-Unemotional Traits in a community sample of young adult males. B A , 35(1), 20–34.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. , 117(3), 817–869.

Chen, Z. (2013). The relationship between adolescent callous-unemotional traits and emotional processing. Master's thesis.

Ci, J. (2009). . Harvard University Press.

Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character. , 21(5), 355–359. https://doi.org/10.1177/0963721412454874

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. , 100(5), 947–966. https://doi.org/10.1037/a0022641

Costa, D. M., & Babcock, J. C. (2008). Articulated thoughts of intimate partner abusive men during anger arousal: Correlates with personality disorder features. , 23(6), 395.

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. A , 111(48), 17320–17325. https://doi.org/10.1073/pnas.1408988111

Dalbert, C. (1999). The world is more just for me than generally: About the personal belief in a just world scale's validity. , 12(2), 79–98.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. , *44*(1), 113–126.

Davis, M. H. (2015). Empathy and prosocial behavior. In D. A. Schroeder & W. G. Graziano (Eds.), (pp. 282–306). Oxford University Press.

Davis, M. H., Mitchell, K. V, Hall, J. A., Lothert, J., Snapp, T., & Meyer, M. (1999). Empathy, expectations, and situational preferences: Personality influences on the decision to participate in volunteer helping behaviors. , *67*(3), 469–503.

De Hooge, I. E. (2019). Improving our understanding of guilt by focusing on its (Inter) personal consequences. In B. Cokelet & C. J. Maley (Eds.), (p. 131). Rowman & Littlefield International.

Di Pierro, R., Sarno, I., Perego, S., Gallucci, M., & Madeddu, F. (2012). Adolescent nonsuicidal self-injury: The effects of personality traits, family relationships and maltreatment on the presence and severity of behaviours. , *21*(9), 511–520.

Dovidio, J. F., Allen, J. L., & Schroeder, D. A. (1990). Specificity of empathy-induced helping: Evidence for altruistic motivation. , *59*(2), 249.

Eisenberg, N., & Valiente, C. (2002). Parenting and children's prosocial and moral development. In M. H. Bornstein (Ed.), *5*, (p. 111). Lawrence Erlbaum Associates Publishers.

Engel, C. (2011). Dictator games: A meta study. , *14*(4), 583–610.

Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Callous-unemotional traits in a community sample of adolescents. *A* , *13*(4), 454–469.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation*. , *114*(3), 817–868.

Fehr, E., & Schmidt, K. M. (2010). On inequity aversion: A reply to Binmore and Shaked. *B* & , *73*(1), 101–108.

FeldmanHall, O., Dalgleish, T., Evans, D., & Mobbs, D. (2015). Empathic concern drives costly altruism. , *105*, 347–356.

Ferguson, T. J., & Stegge, H. (1995). Emotional states and traits in children: The case of guilt and shame. In J. P. Tangney & K. W. Fischer (Eds.), (pp. 174–197). Guilford Press.

Fliessbach, K., Phillipps, C. B., Trautner, P., Schnabel, M., Elger, C. E., Falk, A., & Weber, B. (2012). Neural responses to advantageous and disadvantageous inequity. , *6*, 165.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *B* , *6*(3), 347–369.

Frick, P. J., & Viding, E. (2009). Antisocial behavior from a developmental psychopathology perspective. *D* , *21*(4), 1111–1131. https://doi.org/10.1017/S0954579409990071

Gao, X., Yu, H., Sáez, I., Blue, P. R., Zhu, L., Hsu, M., & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous-and disadvantageous-inequity aversion. *A* , *115*(33), E7680–E7689.

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. , *5*, e11305.

Giner-Sorolla, R., Piazza, J., & Espinosa, P. (2011). What do the TOSCA guilt and shame scales really measure: Affect or action? *D* , *51*(4), 445–450.

Glenn, A. L., & Raine, A. (2014). *: A* (Vol. *1*). NYU Press.

Gong, X., Brazil, I. A., Chang, L. J., & Sanfey, A. G. (2019). Psychopathic traits are related to diminished guilt aversion and reduced trustworthiness during social decision-making. , *9*(1), 7307. https://doi.org/10.1038/s41598-019-43727-0

Gregory, S., Blair, R. J., Simmons, A., Kumari, V., Hodgins, S., & Blackwood, N. (2015). Punishment and psychopathy: A case-control functional MRI investigation of reinforcement learning in violent antisocial personality disordered men. , *2*(2), 153–160.

Güroğlu, B., Will, G.-J., & Crone, E. A. (2014). Neural correlates of advantageous and disadvantageous inequity in sharing decisions. , *9*(9), e107996.

Hayes, A. F. (2013). *C* *A* *: -B A* . New York, NY: Guilford Press.

Hinde, R. A. (2002). *:* . Psychology Press.

Hoffman, L. W. (1991). The influence of the family environment on personality: Accounting for sibling differences. *B* , *110*(2), 187.

Kamau, C., Giner-Sorolla, R., & Zebel, S. (2013). Reconciliation responses, blame, and expressions of guilt or shame. *A* , *43*(SUPPL.2), 287–292. https://doi.org/10.1111/jasp.12048

Kang, C., Zhao, X., Xu, X., Yang, K., & Yang, J. (2001). The Questionnaire of Systemic Family Dynamics: Development, Reliability and Validity. *C* , 92–95.

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., Aucoin, K. J., & Morris, A. S. (2008). Assessing callous–unemotional traits in adolescent offenders: Validation of the Inventory of Callous–Unemotional Traits. , *31*(3), 241–252.

Kimonis, E. R., Kidd, J., Most, S. B., Krynen, A., & Liu, C. (2019). An elusive deficit: Psychopathic personality traits and aberrant attention to emotional stimuli. , *20*(6), 951–964.

Klein, N., & Epley, N. (2016). Maybe holier, but definitely less evil, than you: Bounded self-righteousness in social judgment. , *110*(5), 660.

Kochanska, G. (1991). Socialization and temperament in the development of guilt and conscience. *C D* , *62*(6), 1379–1392.

Kochanska, G. (1997). Mutually responsive orientation between mothers and their young children: Implications for early socialization. *C D* , *68*(1), 94–112.

Koller, M. (2016). Robustlmm: an R package for robust estimation of linear mixed-effects models. *Journal of Statistical Software*, *75*(6), 1–24.

Loehlin, J. C., & Nichols, R. C. (2012). *Heredity, environment, and personality: A study of 850 sets of twins*. University of Texas Press.

López, E. E., Pérez, S. M., Ochoa, G. M., & Ruiz, D. M. (2008). Adolescent aggression: Effects of gender and family and school environments. *Journal of Adolescence*, *31*(4), 433–450.

McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, *1*(2), 1–9.

McAuliffe, K., Chang, L. W., Leimgruber, K. L., Spaulding, R., Blake, P. R., & Santos, L. R. (2015). Capuchin monkeys, Cebus