Title
Advanced Mental State Reasoning in Children and Adults

Author
Kramer, Hannah Jean

Publication Date
2021

Peer reviewed|Thesis/dissertation

Advanced Mental State Reasoning in Children and Adults

By

HANNAH J. KRAMER

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Dr. Kristin Hansen Lagattuta, Chair

Dr. Simona Ghetti

Dr. Alison Ledgerwood

Committee in Charge

2021

Dedication

I dedicate my dissertation to my dad, Seymour Kramer. When I was six years old, my dad and I were sitting on a plane waiting for it to take off. He explained that my 2-year-old sister, Sasha, did not recognize that her mind was separate from everyone else's. I found this hilariously fascinating. He was always my biggest fan, but he also said that he never really understood what I was studying. Little did he know that this conversation was one of my first inspirations for studying beliefs about minds. Among other things, I wish I could have told him this.

Acknowledgements

Portions of this research were supported by the Predoctoral Training Consortium in Affective Science from the National Institute of Mental Health (201302291) to Hannah Kramer, by a National Science Foundation Rapid Response Research Program (1649566) awarded to Dr. Paul Hastings and Dr. Luis Parra, and by UC Davis Internal Grants to Dr. Kristin Hansen Lagattuta. I appreciate the participants who volunteered their time to the studies presented.

Dr. Kristin Hansen Lagattuta has been my advisor, mentor, and (now) friend for nearly 10 years, so any attempt at summarizing what I have learned from her, my admiration, and the many jokes that we have shared will fall short of the true gratitude that I feel.[1] I want to thank Kristin for believing in me as a researcher, but simultaneously questioning my ideas. This feeling of support, combined with a need to clarify my thinking through discussion has made me the scientist I am today. Kristin's fearlessness when studying new topics was one reason I wanted to work with her and part of why I am so sad to leave. I appreciate the opportunities that I had to write alongside Kristin, her instilling the need for precision (e.g., make word counts exactly as requested, ensure paragraphs end at the right-hand margin, avoid repeating words), and her invention of "wondering" drafts. To put this all another way, Kristin made graduate school the "evening of my life" rather than the "life of my evening," and I cannot thank her enough.

From the moment I joined Dr. Alison Ledgerwood's lab as a graduate student affiliate, she welcomed me as if I was one of her own students. I admire Alison's ability to ask insightful questions in the most kind and constructive way. I am grateful to have had the opportunity to learn from her. The faculty and staff at UC Davis have been very supportive; I am especially grateful for the feedback that I have received on my research from Drs. Simona Ghetti, Amanda

[1]Dr. Karen Lara can attest to the many drafts. Also, Kristin knows that I love a good footnote.

Guyer, Paul Hastings (contributed to Chapter 2), Andy Todd, Paul Eastwick, Lindsay Bowman, Jeff Sherman, and Cindy Pickett. I thank Drs. Hyo Gweon and Jamil Zaki for their contributions to Chapter 1, and for their kindness when I did rotations in their labs.

My "co-worker," Dr. Karen Lara, is one of the first people I tell ideas to because of her honesty (but, I am still mad about the snail and the cheetah). I will miss the hours that we spent (not) thinking in the "no thinking zone," but I am not even a "a little worried" about the future of our friendship. I am "very" thankful that I found an amazing collaborator early in my career. Thank you to the research assistants and graduate students that I have learned from, especially Nichole Bouffard, Chris Coleman, Aline da Silva Frost, Dr. Deborah Goldfarb (contributed to Chapter 3), MJ Heise, Courtney Jimenez, Dr. Katie Kennedy, Sam Klein, Dr. Karen Lara (contributed to Chapter 2), Abby Lavine, Dr. Luis Parra (contributed to Chapter 2), Riley Sims, Dr. Jehan Sparks, Dr. Sarah Tashjian (contributed to Chapter 3), Austen Trainor, Andre Wang, and Julia Wu. You each make me so excited about science! Thank you to Sarah Lagattuta for (attempting) to teach me to dance, for the many games she created (e.g., Alpaca Wars, the Piano Game), and for popcorn and ice cream (but, only on Fridays) while watching MAFS.

With Stephen Paolillo I can be my full self. I am thankful for him and the unwavering support that he provides me (thanks to Sarah Smith for introducing us). I would not have found research if I had not observed the passion that that my mom, Laurie Goldsmith, has for her own work as a clinical psychologist. I am fortunate to have a mom who supports me in everything that I do. I can tell her anything and she listens fully. I am thankful to have my sister, Sasha Kramer, as a constant in my life. I am so proud of who she is and I cannot wait to buy her a peacoat. To the animals (especially Zeno, Rice, Oink, Ruby, Otto, Roo, Piper, Nels, Joel, Bailey, Moo, Ziti, Grey, Misty, Socks, and Scout), thank you for the joy.

Table of Contents

List of Tables

List of Figures

Abstract

People navigate the social world by considering the invisible, but ever-present mental lives of themselves and others. I present three papers that push the boundary of what we know about mental state reasoning by testing children's and adults' knowledge of complex aspects of the mind. In Chapter 1, I tested whether children and adults incorporate temporal markers into their mental state reasoning. Eight- to 10-year-olds and adults estimated the time duration of emotions (e.g., feeling sad), desires (e.g., wanting strawberries), and preferences (e.g., liking strawberries). I found that children and adults hold lay theories about the time course of mental states: They viewed preferences as longer lasting than emotions and desires, expressed confidence in their duration judgments, and provided internally consistent responses that were reliable across a 1-week delay. In Chapter 2, I addressed adults' beliefs about emotional responses to events as they unfold over time. LGBTQ-Latinx, LGBTQ-White, Straight-Latinx, and Straight-White participants integrated the valence (positive versus negative) and timing of events (first versus last in a sequence) into their emotion ratings. Across demographic groups, adults expected past negative events to shape reactions to later positive events, but prior positive events to have little influence over responses to subsequent negative events. Finally, in Chapter 3, I measured children's and adults' inferences about the abilities, preferences, and traits of a novel social group. I discovered that participants used their knowledge of one group to make inferences about an unknown group in the form of a *dichotomizing heuristic*—what is true of one group must not be true of another group. Adults widely applied this cognitive shortcut when inferring benign (e.g., liking apples), novel (e.g., playing daxes), and even evaluative group characteristics (e.g., being smart). The common thread connecting these papers is an aim to move beyond thinking about mental states as isolated experiences to integrate the broader contexts in which they occur.

Chapter 1

This Too Shall Pass, But When? Children's and Adults' Beliefs about the Time Duration of Emotions, Desires, and Preferences

Abstract

Children and adults (Study 1: *N* = 50 undergraduates; 4% African American, 38% Asian, 2% Hawaiian/Pacific Islander, 24% Hispanic/Latino, 18% White, 7% another race or ethnicity, and 7% Multiethnic or Multiracial; 64% female, 36% male; Study 2: *N* = 112 8- to 10-year-olds and undergraduates; 25% Asian, 12% Hispanic/Latino, 44% White, 4% another rate or ethnicity, 15% Multiethnic or Multiracial; 50% female; 50% male) estimated how long emotions (e.g., sadness), desires (e.g., wanting milk), and preferences (e.g., liking milk) last (seconds, minutes, hours, days, weeks, months, years, decades, a lifetime or longer). Adults rated preferences as longer lasting than emotions and desires. They also judged that self-conscious emotions (i.e., shame, pride) were more stable than basic emotions (e.g., mad, happy). Although children also distinguished among the broad mental state categories (preferences > emotions, desires in duration), they did not differentiate between self-conscious and basic emotions. Both children and adults provided internally consistent responses, exhibited good test-retest reliability, and were confident in their duration judgments.

Keywords: theory of mind, social cognition, emotion, temporal reasoning, middle childhood

This Too Shall Pass, But When?

Children's and Adults' Beliefs about the Duration of Emotions, Desires, and Preferences

Imagine that you have three friends, and they each provide you with a piece of information. One says, "I like to hike" another says, "I want milk," and the third says, "I feel mad." To what extent do each of these statements provide valuable information beyond the moment of disclosure? Although none of your friends mentioned temporal duration, it may be reasonable to have expectations about how relevant each of these descriptions will be at later points in time. For example, you could plausibly anticipate that your friend who "likes to hike" may accept your invitation to hike next week or even many years into the future, whereas your friend who "wants milk" now may not want it even later that day, and your friend who currently "feels mad" may experience varying emotions before you call her back to check-in that evening. Indeed, intuitions about the time course of different mental states may help guide interpersonal interactions as well as facilitate regulation and understanding of one's own mental states.

In the current research, we explored 8- to 10-year-olds' and adults' beliefs about whether mental state terms signal duration information; for example, do people assume that liking lasts longer than wanting? Understanding how children and adults think about mental states through time is critical for enriching scientific understanding about the development of theory of mind—knowledge about one's own and others' mental states (Wellman, 2014)—as well as for ultimately assessing the potential impact of such beliefs on behavior and mental health (e.g., individuals who believe negative emotions are long lasting may be at greater risk for depression; beliefs about the time duration of desires may impact resistance to temptation). Despite the importance of children's and adults' judgments about the time course of mental experiences, little work has tested just how long people expect current mental states to last into the future.

Thus, across two studies, we assessed 8- to 10-year-olds' and adults' lay theories about the time duration of emotions (what someone *feels*), desires (what someone *wants*), and preferences (what someone *likes*). Although measuring beliefs about the duration of many types of mental states is worthy of study, we narrowed our focus to this subset of mental states that have an affective or evaluative component (i.e., emotions, desires, preferences). As will be reviewed, children and adults have difficulty setting aside these kinds of highly salient present states when predicting the future. Such work begs the question as to whether children and adults hold consistent beliefs about the duration of these states when they are not currently experiencing them.

What makes measuring beliefs about the duration of mental states particularly intriguing is that although it is possible that children and adults view mental state terms (e.g., feel, want, like) as having temporal connotations, formal definitions do not actually contain this information (e.g., Merriam-Webster; American Dictionary of the English Language; Dictionary.com). Thus, any consistencies among people's belief systems surrounding the duration of mental states likely would have been learned through informal, but collective processes. This question, then, is especially interesting and important to investigate from a development perspective because if temporal information is not explicitly part of the definition of these mental state terms, then it may take children some time to appreciate the temporal aspects of varying mental states.

From 2 to 10 years of age as well as between childhood and adulthood, individuals increasingly appreciate that emotions are multiplicatively determined: They can be caused by current situations, mental states, and past experiences (for reviews see Harris, 2010; Kramer & Lagattuta, in press; Lagattuta, 2014; Lagattuta & Kramer, 2021; Lagattuta et al., 2015). Thus, children and adults recognize that several factors can lead emotions to change. Less work, however, has directly measured children's and adults' beliefs about how long an emotion

typically lasts. For example, Roberts and Gelman (2016) demonstrated that Black 5- to 6-year-olds (but not White 5- to 6-year-olds), White 9- to 10-year-olds, and White adults expect a person's emotional expressions (versus their race) to be less stable from childhood to adulthood (e.g., they thought it was more likely for a happy white child to grow up to be an angry White adult than a happy Black adult). Relatedly, children as young as 9 years do not expect an emotional experience (i.e., feeling grumpy or nervous) to remain stable from childhood to adulthood (Brandone & Klimek, 2018). In other work, 4- to 6-year-olds were told about a positive or negative event and were asked to determine how the character would feel at different points in time (e.g., later in the morning, after lunch). Six-year-olds thought that emotional reactions would take time to wane, whereas 4-year-olds predicted that emotions would change more quickly (Harris et al., 1985). Lagattuta et al. (1997) also documented improvements between the ages of 4 and 6 in recognition that the re-instatement of negative emotions depends on the focus of a person's thoughts. Finally, although adults expect events (e.g., a breakup) to continue to shape their emotional wellbeing at future time points (e.g., three months later), individuals actually recover from emotional episodes in much shorter time periods than they initially forecast (see Wilson & Gilbert, 2005 for a review).

Like emotions, concepts of desires emerge early in development, but show continued improvement through middle childhood and into adulthood. For example, children's early knowledge that fulfilling desires feels good makes it difficult for them to appreciate that breaking a rule to fulfill a desire can actually feel bad (Arsenio et al., 2006; Barden et al., 1980; Lagattuta, 2005, 2008, 2017; Nunner-Winkler, & Sodian, 1988; Weller & Lagattuta, 2013, 2014). Similarly, although toddlers begin to appreciate that people will diverge in their desires by 18 months of age (Repacholi & Gopnik, 1997; but see Ruffman et al., 2017), even 4- to 5-year-olds

assume common ground in tastes unless they are given explicit counter information (Atance et al., 2010; Cassidy et al., 2005). More pertinent to the current research, children and adults fail to appreciate that their present desires will not carry forward in time. For example, 3- to 13-year-olds induced to feel thirsty overwhelmingly predict that they will want water (more than free pretzels) the following day, not realizing that their current thirst will subside in the interim (Atance & Meltzoff, 2006; Kramer et al., 2017; Mahy, 2015; Mahy et al., 2014; Mazachowsky et al., 2019). Similar biases have also been documented when adults forecast what they will want for the future; they, too are tethered to the present when imagining their future desires (Gilbert et al., 2002; Loewenstein et al., 2003; Read & Van Leeuwen, 1998). Indeed, Kramer et al. (2017) documented lifespan continuity (from preschool to early adulthood) in an overreliance on current states when predicting future desires.

Children and adults use several factors to infer people's preferences. They can infer what someone likes by comparing that person's choice to the probability of that choice occurring at random (e.g., they more often judge that a person likes frogs if they select a frog out of mostly duck toys versus a frog out of mostly frog toys; Choi et al., 2018; Gweon et al., 2010; Kushnir et al., 2010; Ma & Xu 2011; Pesowski et al., 2016; Wellman et al., 2016). Moreover, if children and adults learn what one member of a group prefers, they assume that the preference extends to other group members (Diesendruck et al., 2015; Goldfarb et al., 2017; Kalish & Lawson, 2008). Although determining what others like develops early, estimating the stability of preferences proves challenging even for adults. Decades of research in psychology and economics suggests that people's preferences fluctuate over time (Amir & Levav, 2008; Tversky et al., 1988) and are susceptible to changes in context (Ledgerwood & Boydstun, 2014; Levin & Gaeth, 1988). Despite the shifting nature of preferences, preschoolers view their own likes and dislikes as

stable over time (Belanger et al., 2014), with even adults also expecting future continuity in their current preferences (Bauckham et al., 2017; Quoidbach et al., 2013; Renoult et al., 2016).

Present Research

In all of the paradigms reviewed, researchers have asked children and adults to assess whether an emotion, desire, or preference will still be present at a later point in time. Previous studies have not tested how long children and adults believe that an individual will *continuously* feel, want, or like something. Moreover, because most research on developmental social cognition has examined children's and adults' knowledge about a particular mental state (e.g., children's understanding of emotion, but see Lagattuta et al., 2016), it remains unclear how children and adults differentiate among various mental states; that is, what do they think makes a desire different from an emotion or a preference? This distinction is particularly interesting for differences between desires and preferences. Indeed, although we have split our discussion of the two affective states (desires vs. preferences) into separate sections, these decisions were somewhat arbitrary in that it is not always clear whether researchers are measuring beliefs about preferences (i.e., likes) or desires (i.e., wants), as sometimes they use these terms interchangeably. We propose that one important feature that children and adults may use to distinguish among affective states (emotions, desires, and preferences) is their time duration. To test this, we developed a new measure to assess beliefs about how long mental states last.

Participants were told about a person who has a mental state, and they judged how long that experience would last (e.g., "Sam wants milk. How long will Sam want milk?"). In addition to providing temporal estimates, participants also indicated how confident they were in their judgments. Study 1 focused on young adults. In Study 2, we tested 8- to 10-year-olds and adults to assess age-related differences and similarities in beliefs about mental states across time.

Study 1

The goal of Study 1 was to create a new measure to assess beliefs about mental states across time. This study included only adults so that we could determine whether they were able to use the measure appropriately prior to administering it to children. In this study, we analyzed whether adults would provide differentiated temporal estimates for emotions, desires, and preferences. Although this was an exploratory study, our intuition was that adults would rate emotions and desires as more temporary than preferences.

Method

Participants

Fifty adults participated for course credit. Because this was the first study to examine people's beliefs about the comparative duration of mental states, we could not conduct a power analysis. We chose to collect 50 participants based on rule of thumb. After exclusions based on *a priori* criteria (see below) the final sample included 45 adults (M = 20.98 years, SD = 2.53 years; range: 18.05 years to 31.57 years; 29 females, 16 males). For our main analysis in this study, a 5 (mental state: temporary control, stable control, emotion, desire, preference) repeated measures ANOVA, this sample size gave us 80% power to detect an effect size $f = .23$ (i.e., a $\eta_p^2 \approx .05$). The final sample was diverse in ethnicity, parental income, and parental education: Participants were 4% African American, 38% Asian, 2% Hawaiian/Pacific Islander, 24% Hispanic/Latino, 18% White, 7% Multiethnic or Multiracial, and 7% another race or ethnicity. Thirty percent of participants had parents who made less than $50,000, 30% had parents who made between $50,001 and $100,000, and 28% had parents who made more than $100,000; 9% of the sample did not report this information. Thirteen percent of participants' parents did not have a high school degree, 16% had a high school degree, 15% had an Associate's degree or attended some

college, 24% had a college degree, and 31% had completed some graduate work or had a graduate/professional degree. For participating, adults received course credit. Data were collected from April 2017 to May 2017. This project was approved by the Internal Review Board at the [BLINDED]: Protocol name: [BLINDED]; Protocol Number: 1031991.

Mental States Across Time (MAT)

After providing informed consent and completing a demographics questionnaire, participants responded to the new *Mental States Across Time Task* (MAT) administered via Qualtrics Survey Software (Qualtrics, 2021). For each trial, participants were told about someone currently experiencing a mental state (e.g., feeling happy; wanting milk, liking to hike), and they rated how long that mental state would last (e.g., "Something happened that made Casey feel worried. How long will Casey feel this?"). Names were gender neutral and changed for each statement. Participants judged duration in a two-step process: First, they selected a temporal unit (seconds, minutes, hours, days, weeks, months, years, decades, a person's lifetime or longer),[2] and then they determined the precise quantity (e.g., how many seconds). After choosing the time unit (e.g., seconds), but before providing the exact number (e.g., 15 seconds), participants also rated how certain they were about their time duration unit on a sliding scale (from 0 to 100: *not at all sure* to *completely sure*). The referents for desire and preference trials were identical (i.e., wanting/liking to hike; wanting/liking milk) so that any differences between them could only be explained by distinctions in mental state term (want versus like) versus referent type. Although we focused on emotions (12 trials), desires (2 trials), and preferences (2 trials), participants judged time durations for several additional mental states and characteristics (e.g., knowledge states, abilities, race, gender; see Table A.1; 42 total trials).

[2]We provided participants with instructions of what we meant by each of the time units (e.g., by seconds we mean less than 60 seconds).

We included multiple attention checks throughout the measure. During pilot testing, several participants mentioned that they wanted to know the age of the people in each of the scenarios. For this reason, at the beginning of the task we informed all participants that they should assume that the character in each scenario was 20 years old. Immediately following this information (and with it still visibly present), we asked participants for the age of each of the characters. Any participant who got this question wrong was removed from analyses (n = 2).[3] As well, participants were asked to determine how long they expected each of the following to last: an ice cube, a rock, a sunburn, and eye color. We included these control trials to ensure that participants understood the temporal scales. These items were randomized with the rest of the questions. We decided (*a priori*) to exclude participants if they said that an ice cube lasts longer than a rock (n = 3) or if they said that a sunburn lasts longer than a person's eye color (n = 0).

Results

Results are divided into two overarching sections: duration estimates and certainty. We examined whether adults provide differentiated duration estimates and certainty ratings by type of mental state (temporary control vs. emotions vs. desires vs. preferences vs. stable control). In addition to the overarching mental state categories, we also examined whether adults provided distinct duration estimates based on emotion type (happy, sad, proud, ashamed, embarrassed, mad, afraid, excited, worried, surprised, startled, okay). We conducted analyses in RStudio (Version 3.6.1; R Core Team, 2019).

Duration Estimates

Duration Estimates by Mental State. A 5 (mental state: temporary control, emotion, desire, preference, stable control) repeated measures analysis of variance (ANOVA) on duration

[3]One additional participant skipped this question, but they got a memory check question about the characters' ages correct at the end of the study, so this participant was included in analyses.

estimates resulted in a main effect for mental state, $F(4, 176) = 599.85$, $p < .001$, $\eta_p^2 = .93$ (see *Table 1* and *Figure 1*). Post-hoc comparisons (Tukey's HSD) indicated that adults judged emotions as more temporary than desires ($p = .017$), which were judged to be less stable than preferences ($p < .001$). In addition, adults judged that emotions and desires were as fleeting as the temporary controls ($ps > .110$), and preferences were as long lasting as the stable controls ($p = .721$). In other words, adults expected the temporary controls (i.e., an ice cube, sunburn), emotions, and desires, to last between hours and days, but they anticipated that preferences and stable controls (i.e., a rock, a person's eye color) would last between decades and a person's lifetime. This pattern held at the level of the individual: 96% of participants judged that preferences would last longer than both desires and emotion (62% of participants also judged that desires last longer than emotions).[4]

[4]We also examined emotions, desires, and preferences using the discrete duration estimates (i.e., the answer to the question how many seconds/minutes/hours/days/weeks/months/years/decades) as the dependent variable. We converted estimates into hours (if a participant provided a discrete estimate outside of the defined range for that unit, we used the largest number of that unit; for a person's lifetime or longer, we used 80 years). This analysis yielded similar results except that emotion and desire estimates were equivalent ($p = .817$).

Table 1. Chapter 1: Study 1 Means [95% Confidence Intervals] of Duration Estimates and Certainty Ratings by Item

Item	Duration	Certainty
Ice Cube: Ellis see that there is an ice cube on a table.	2.07 [1.95, 2.18]	86.07 [80.28, 91.85]
Sunburn: Peyton has a sunburn.	4.31 [4.14, 4.48]	81.53 [75.16, 87.91]
Temporary Control	3.19 [3.08, 3.30]	83.80 [78.77, 88.83]
Rock: Jordan is on a hike and sees there is a rock.	8.42 [8.04, 8.80]	91.56 [86.15, 96.96]
Eye Color: Max has brown eyes.	9.00 [no variability]	98.09 [96.58, 99.60]
Stable Control	8.71 [8.52, 8.90]	94.82 [91.81, 97.84]
Happy: Something happened that made Frances feel happy.	3.11 [2.81, 3.41]	64.58 [57.59, 71.56]
Sad: Something happened that made Ray feel sad.	3.47 [3.09, 3.84]	61.80 [55.14, 68.46]
Proud: Something happened that made Quinn feel proud.	4.40 [3.75, 5.05]	65.98 [58.80, 73.15]
Ashamed: Something happened that made Hayden feel ashamed.	3.98 [3.51, 4.45]	60.53 [52.54, 68.52]
Embarrassed: Something happened that made Chris feel embarrassed.	3.49 [2.94, 4.03]	63.64 [57.49, 69.80]
Mad: Something happened that made Rory feel mad.	3.09 [2.77, 3.41]	65.33 [58.05, 72.62]
Afraid: Something happened that made Brett feel afraid.	2.93 [2.53, 3.34]	59.09 [51.03, 67.15]
Excited: Something happened that made Drew feel excited.	2.76 [2.48, 3.04]	62.96 [55.47, 70.44]
Worried: Something happened that made Casey feel worried.	3.24 [3.00, 3.49]	63.20 [56.05, 70.35]
Surprised: Something happened that made Aubrey feel surprised.	1.73 [1.50, 1.97]	71.87 [65.51, 78.23]
Startled: Something happened that made Jo feel startled.	1.84 [1.56, 2.13]	68.04 [60.61, 75.48]
Okay: Something happened that made Jessie feel okay.	3.22 [2.87, 3.57]	59.76 [52.16, 67.35]
Emotion (Feels)	3.07 [2.87, 3.27]	63.90 [58.43, 69.36]
Desire Milk: Addison wants milk.	2.98 [2.39, 3.57]	63.44 [55.82, 71.07]
Desire Hike: Morgan wants to go hiking.	4.22 [3.87, 4.57]	61.58 [54.15, 69.01]
Desire (Want)	3.60 [3.19, 4.01]	62.51 [55.66, 69.36]
Preference Milk: Robin likes milk.	8.64 [8.45, 8.84]	77.84 [70.89, 84.79]
Preference Hike: Corey likes to go hiking.	8.36 [8.11, 8.60]	78.07 [71.23, 84.90]
Preference (Like)	8.50 [8.32, 8.68]	77.96 [72.44, 83.47]

Note. Duration estimates: 1 = seconds, 2 = minutes, 3 = hours, 4 = days, 5 = weeks, 6 = months, 7 = years, 8 = decades, 9 = a person's lifetime or longer. Certainty ratings: 0 = "not at all sure" to 100 = "completely sure." For ten participants, the wording for the Desire Hike and Preference Hike questions was worded differently (Desire: "Morgan wants to hike" Preference: "Corey likes to hike.

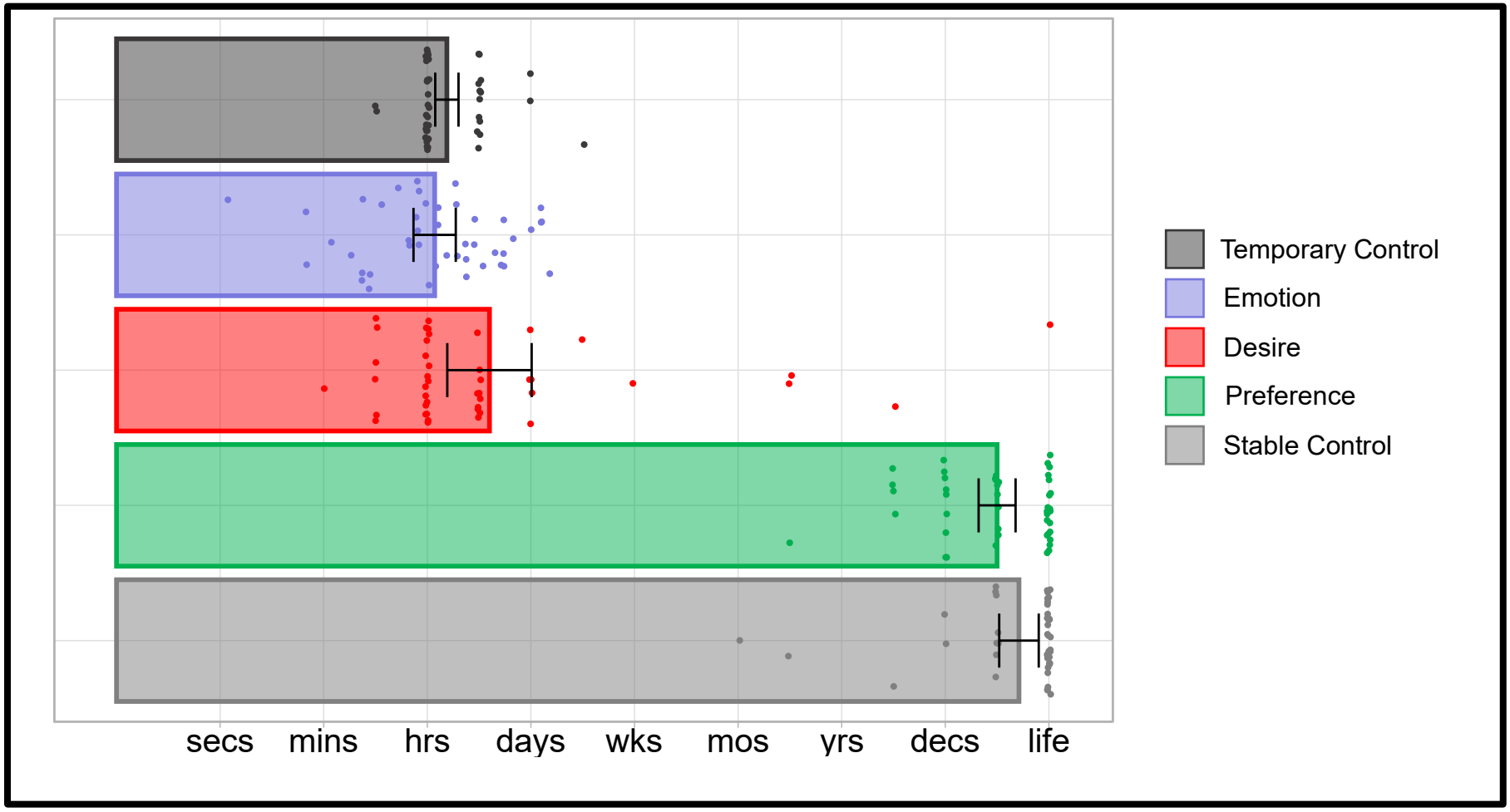

Figure 1. Chapter 1: Study 1 Adults' duration estimates by mental state type. Bar = mean; error bar = 95% confidence interval; dot = jittered individual. Adults differentiated mental states by their duration: stable control = preference > desire > emotion = temporary control.

Duration Estimates by Discrete Emotion. A 12 (emotion: happy, sad, proud, ashamed, embarrassed, mad, afraid, excited, worried, surprised, startled, okay) repeated measures ANOVA on duration estimates (1 = seconds, 2 = minutes, 3 = hours, 4 = days, 5 = weeks, 6 = months, 7 = years, 8 = decades, 9 = a lifetime) resulted in a main effect for emotion $(11, 484) = 20.08$, $p < .001$, $\eta_p^2 = .31$ (see *Table 1 and Figure 2*). Post-hoc comparisons (Tukey's HSD), showed that participants rated surprised and startled as the most fleeting emotions (*p*s < .009). Adults also rated proud and ashamed as the most stable emotions (*p*s < .018), except that ashamed was no different from sad ($p = .603$), embarrassed ($p = .669$), worried ($p = .097$), or okay ($p = .075$). All other emotions were judged to last a comparable amount of time (*p*s > .097).

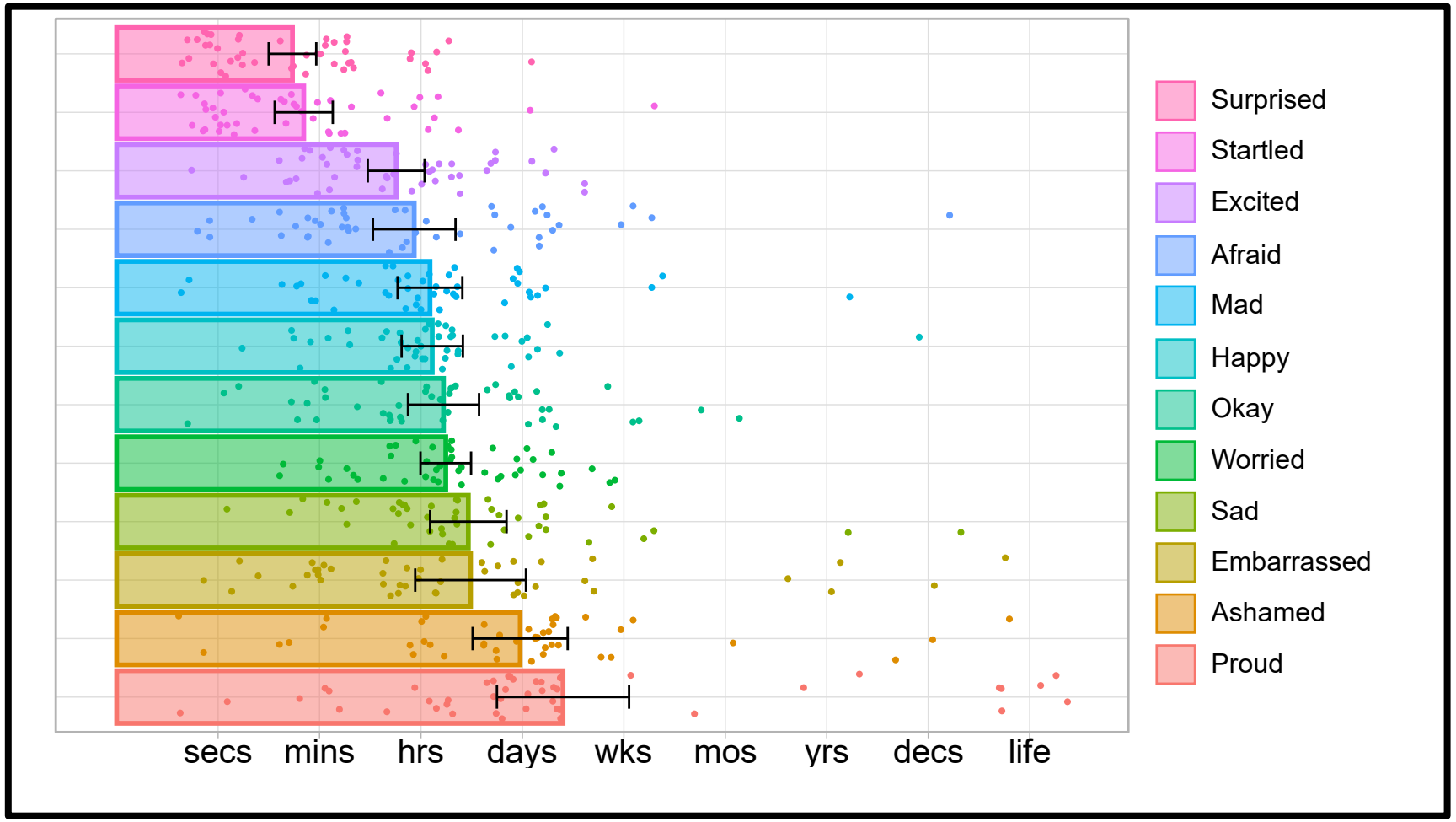

Figure 2. Chapter 1: Study 1 Adults' duration estimates by emotion type. Bar = mean; error bar = 95% confidence interval; dot = jittered individual. Adults differentiated discrete emotions states by their duration: surprised and started were the most fleeting and ashamed and proud were the most stable.

Consistency in Duration Estimates by Mental State. We examined whether adults provided internally consistent duration judgments for each of the 12 emotions: The internal reliability was adequate (α = .79, 95% CI [.70, .88]). Participants' judgments about desire duration across both trials significantly correlated (r[43] = .47, 95% CI[.20, .67], p = .001), and the two preference trials were marginally correlated (r[43] = .28, 95% CI[-.01, .53], p = .059). Finally, we tested whether participants' mental state duration estimates correlated with each other. None of these reached conventional levels of significance (emotion vs. desire: r[43] = -.04, 95% CI [-.33, .25], p = .781; emotion vs. preference: r[43] = -.07, 95% CI [-.36, .22] p =.628; desire vs. preference: r[43] = -.22, 95% CI [-.48, .08], p = .148). This pattern of stronger response consistency within versus between mental states suggests that it is not the case that some participants simply think that mental states are short lived whereas other participants think that mental states are long-lasting; they distinguish duration by mental state type.

Certainty Ratings

Certainty Ratings by Mental State. Participants were confident in their duration estimates (compared to the scale midpoint of 50: temporary control: t[44] = 13.54, p < .001, d = 2.02; emotion: t[44] = 5.12, p < .001, d = 0.76; desire: t[44] = 3.68, p < .001, d = 0.55; preference: t[44] = 10.21, p < .001, d = 1.52; stable control: t[44] = 29.98, p < .001, d = 4.47). To explore differences by mental state, we conducted a 5 (mental state) repeated measures ANOVA on average certainty ratings. Mental state was a within-subjects factor. This analysis resulted in a main effect for mental state, F(4, 176) = 45.60, p < .001, η_p^2 = .51 (*Table 1*). Post-hoc comparisons (Tukey's HSD), showed that adults were more confident in their duration estimates for preferences than they were for emotions or desires (ps < .001; desire = emotion, p = .989). Participants were most confident in duration estimates for stable controls (ps < .002), and

they were also more certain about duration estimates for temporary controls than they were for rating how long emotions and desires last (*p*s < .001; temporary control = preference, *p* = .250).

Consistency in Certainty Ratings by Mental State. We examined whether the confidence ratings for the 12 emotions were internally consistent: The internal reliability was excellent (α = .93, 95% CI [.90, .96]). Participants' confidence ratings for desire duration estimates were significantly correlated (*r*[43] = .66, 95% CI[.45, .80], *p* = .001) and their preference certainty ratings were marginally (but non-significantly) correlated (*r*[43] = .28, 95% CI[-.01, .53], *p* = .062). Finally, we tested the extent to which participants' mental state duration estimates were correlated with each other. All relations were significant (emotion vs. desire: *r*[43] = .88, 95% CI [.79, .93], *p* < .001; emotion vs. preference: *r*[43] = .57, 95% CI [.33, .74] *p* < .001; desire vs. preference: *r*[43] = .41, 95% CI [.13, .63], *p* < .001, indicating that some participants are overall more confident in their duration estimates than are other participants, but individual-level confidence is not fully dependent on mental state type.

Discussion

Study 1 provided evidence that the new MAT assessment can measure adults' beliefs about the duration of mental states. The majority of adults used the scale appropriately on control trials that we determined had "correct" responses (i.e., that a rock will last longer than an ice cube; that eye color will last longer than a sunburn). More importantly, adults provided differentiated duration estimates based on mental state type. That is, they expected emotions to be more temporary than desires, and for desires to be more fleeting than preferences. We also documented that adults differentiate discrete emotions by duration: They anticipated surprised and startled to be the most temporary and self-conscious emotions (shame and pride, but not embarrassment) to be particularly long-lasting. Further, this study provided initial evidence that

beliefs about the duration of mental states may be theory-like. That is, we documented within mental state but not between mental state correlations in duration estimates (e.g., wanting milk and wanting to hike was correlated, but wanting and liking judgments did not correlate). Further indication that responses reflect solidified beliefs came from decision confidence ratings: Adults provided consistently high certainty ratings when judging the duration of all mental states, and they were especially confident about the long-lasting duration of preferences. Thus, this study reveals that this new scale taps into people's beliefs about mental states across time.

Study 2

The goal of Study 2 was to replicate and extend Study 1, as well as to incorporate a developmental perspective by including 8- to 10-year-olds. In addition, we aimed to test the reliability of children's and adults' mental state duration estimates by having participants complete the same measure twice about a week apart. We further widened the range of desire and preference scenarios, and balanced them with the number of emotion judgments that participants provided (i.e., eight trials for each type). We tested a relatively older child age group for both practical and theoretical reasons. First, because 8- to 10-year-olds can read (Leppänen et al., 2008), we could administer the same computerized task to children and adults. In addition, it is not until about 7 years that children exhibit knowledge of temporal language (e.g., minutes, hours) equivalent to adults (Tillman & Barner, 2015). Thus, developmental differences in a younger sample could be due to lack of temporal language knowledge, not because of differences in beliefs about the duration of mental states. To ensure that the recruited child sample also applied temporal language appropriately, we included a temporal reasoning measure.

Although research on theory of mind has focused on children under 5 years of age (for reviews, Wellman, 2014; Wellman et al., 2001), there is growing interest in measuring theory of

mind in older age groups (Bernstein, 2018; Hughes, 2016; Kramer & Lagattuta, in press; Lagattuta & Kramer, 2021; Lagattuta et al., 2015). Importantly, several aspects of mental state reasoning that show protracted development require awareness of the impact of mental states over time. For example, children's appreciation that prior expectations shape future emotions is still not fully formed by 8 to 10 years of age (Lara et al., in press; Lara et al., 2019). In addition, although 8- to 10-year-olds recognize that a person's life history influences their future-oriented mental states, and can even generalize to similar contexts, this awareness is more robust in adults versus 8- to 10-year-olds (Kramer et al., 2020; Lagattuta & Kramer, 2021; Lagattuta & Sayfan, 2013). Furthermore, 9-year-olds have more difficulty than adults anticipating when they might experience regret caused by counterfactual thinking (Guttentag & Ferrell, 2008; McCormack & Feeney, 2015). Thus, it is possible that 8- to 10-year-olds' beliefs about the time duration of different mental states may be less defined than that of adults. Of course, because by 8 to 10 years children will have had several experiences with their own and other people's emotions, desires, and preferences moving through time, it is also possible that they have formed solid intuitions about the duration of mental experiences that approximate those of adults.

The distinction that adults made in Study 1 between desires and preferences is intriguing especially given that the only difference between statements was the mental state term (i.e., want vs. like). One difference between wants and likes that could potentially explain the discrepancy is that participants may have interpreted the "want" trials to be inquiring about how long it would take to fulfill the desire, not how long the desire lasts if unfulfilled. In contrast, a preference (or what someone likes) should presumably be less affected by in-the-moment fulfillment versus blockage. Thus, in Study 2, we explored whether providing information about fulfillment or blockage influenced participants' judgments about the relative duration of desires versus

preferences. In addition, because an emotion must derive from an event or else it could be perceived as a mood (Frijda et al., 1991), we tested whether participants' emotion duration estimates depend on the mention of a precipitating cause. That is, we explored whether varying the presence (e.g., "Something happened that made Sam feel happy. How long will Sam feel happy?") versus absence of an instigating event (e.g., "Sam feels happy. How long will Sam feel happy?") would influence participants' emotion duration ratings.

We had three primary hypotheses regarding adults' judgments. First, we expected that adults would provide differentiated temporal estimates for emotions, desires, and preferences (emotions < desires < preferences in time duration; replicating Study 1). Second, we anticipated that adults would rate self-conscious emotions (proud, ashamed) as especially stable and surprised and startled as particularly temporary (replicating Study 1). In addition, we expected adults' duration estimates to take the form of an intuitive theory: That is, we anticipated that their duration estimates would be internally consistent for each mental state (replicating Study 1); that there would be moderate to high test-retest reliability for each mental state (new to Study 2); and, that adults would be confident in their duration estimates (replicating Study 1). As a secondary hypothesis new to Study 2, we expected children to show the same patterns as adults but to a weaker extent. We hypothesized that children would exhibit less differentiation in their duration estimates among the three mental states, weaker internal consistency, lower test-retest reliability, and lower confidence in their duration estimates. We explored how fulfillment versus blockage information shaped judgments about the duration of desires and preferences as well as how the exclusion of an emotion cause influence estimates of the duration of emotions.

Method

We pre-registered our hypotheses, sample size (including planned exclusions), method,

and analyses (https://osf.io/7jdb5/?view_only=ae4c3d17f83046caa181e2fa42a7f2d3).

Participants

Participants included 112 8- to 10-year-olds (n = 54; *M* = 9.47 years; *SD* = 0.94 years 27 females, 27 males) and adults (n = 58; *M* = 20.62 years; *SD* = 1.66 years, 29 females, 29 males). We conducted an *a priori* power analysis in G*Power to detect a 2 (age: 8- to 10-year-olds, adults) x 5 (mental state: temporary control, emotion, desire, preference, stable control) interaction with a small to medium effect size ($\eta_p^2 \approx .03$), our key analysis of interest. To obtain 80% power, the analysis determined that we needed 50 participants per age group. Because we ran this study along with other experiments that required 54 per age group, *N* = 108 was our target sample size. Our ultimate sample exceeded this target because the measures included in this manuscript were part of a larger, three-visit project on children's and adults' social reasoning. When a participant did not complete all three visits, they were replaced, but we still used their data in this manuscript if they had completed the relevant measures. After exclusions (see below) the final sample for analyses was *N* = 108 (52 8- to 10-year-olds and 56 adults).

Child participants were 6% Asian, 8% Hispanic/Latino, 58% White, 23% Multiethnic or Multiracial, and 4% another race or ethnicity; 2% did not report this information. Six percent of child participants had parents who made $50,000 or less, 20% had parents who made between $50,001 and $100,000, and 71% had parents who made more than $100,001; 4% of the sample did not report this information. Ten percent of child participants' parents had an Associate's degree or attended some college, 37% had a college degree, and 54% had completed some graduate work or had a graduate/professional degree. Adult participants were 43% Asian, 16% Hispanic/Latino, 30% White, 7% Multiethnic or Multiracial, and 4% another race or ethnicity. Twenty-four percent of adult participants had parents who made $50,000 or less, 28% had

parents who made between $50,001 and $100,000, and 47% had parents who made more than $100,001; 4% of the sample did not report this information. Eleven percent of adults' parents did not have a high school degree, 14% had a high school degree, 15% had an Associate's degree or at attended some college, 41% had a college degree, and 21% had a graduate degree. For participating, children received $15.00 and two small prizes; adults received course credit. Data were collected from February 2019 to December 2019. This project was approved by the Internal Review Board at the [BLINDED]: Protocol name: [BLINDED]; Protocol Number: 1032366.

Materials and Procedures

Temporal Reasoning. During their first visit to the lab, prior to completing the updated Mental States Across Time Task (MAT, described in detail below), participants completed a *temporal reasoning measure* adapted from Tillman and Barner (2015). First, participants determined which of two units of time was longer (*time comparison questions;* e.g., which is longer seconds or minutes; 11 questions; order of questions randomized). Second, participants transformed time units into different time units (*time transformation questions;* e.g., how many seconds are there in a minute; seven questions; order of questions randomized). Finally, participants ranked 8 units of time from shortest to longest (*time ranking questions;* seconds, minutes, hours, days, weeks, months, years, decades; order of initial order randomized).

Coding and Scoring. For each time comparison question, participants were correct if they selected the longer duration (= 1) and incorrect if they selected the shorter duration (= 0). For time transformation questions, participants were correct if they provided the correct number (= 1) and incorrect if they provided the incorrect number (= 0). For time ranking questions, each time unit was treated as a separate question; participants were correct if they put a given time unit in the correct position (= 1) and incorrect if they put a given time unit in the incorrect

position (= 0). Based on these scores, we calculated a *temporal reasoning score* which equaled the proportion of temporal reasoning questions correct (out of 26 questions). When participants failed to answer a question, their proportion score was calculated out of the total number of questions they answered (1 adult participant answered correctly on all of the comparison questions, but then skipped the remaining questions). Participants who were outliers (3 SDs below the mean) for their age group (8/10 Years: 3 SDs below the mean = .71; adults: 3 SDs below the mean = .89) were excluded (2 8- to 10-year-olds; 1 adult; $N = 109$).

Mental States Across Time (MAT). After completing the temporal reasoning measure, participants responded to an updated MAT. Before beginning, child participants were instructed to imagine that the person in each situation was 10 years old and adult participants were told to imagine that the person in each situation was 20 years old. We chose this age group matching method so that participants could draw from their own personal experiences as opposed to trying to infer the mental states of someone much older or younger. Immediately following this information (and with it still visibly present), participants were asked for the age of the characters. Participants were not able to move on until they got this question correct. We also explained to participants what we meant by each of the time units (e.g., "By seconds, we mean less than 60 seconds"). The vocabulary and reading level of the MAT was at first-grade reading level as assessed via (Flesch Reading Grade Level; Microsoft Word, 2019).

Next, participants were trained to use a certainty scale (from 0 to 100 with five written labels evenly distributed across the scale: *very unsure*, *kind of unsure*, *neither unsure nor sure*, *kind of sure*, *very sure*). First, an experimenter labeled each of the written labels and showed participants that they could actually click anywhere on the scale. Participants then practiced using the scale: The experimenter asked the participant to click in four places (e.g., "Where

would you click if you were *very sure*?"). The four labels for training included one *very*, *one kind of, neither unsure nor sure*, and one in between two points (e.g., "Where would you click if you were between *kind of sure* and *very sure*?"). Participants were not able to move on until they reached 100% accuracy on this training.[5]

Following this scale training, participants completed the MAT. On each trial, children and adults determined how long an emotion, desire, preference, temporary control, or stable control would last (e.g., "Something happened that made Sam feel happy. How long will Sam feel happy?"; "Alex wants milk. How long will Alex want milk?"). The person's name changed for each statement, and all names were gender neutral. The time duration options included: seconds, minutes, hours, days, weeks, months, years, decades, a person's lifetime or longer. In particular, participants reasoned about eight emotions (startled, surprised, mad, sad, worried, happy, proud, ashamed), eight desires (milk, juice, pretzels, strawberries, juice, read, hike, dance, cook), eight preferences (milk, juice, pretzels, strawberries, juice, read, hike, dance, cook), eight temporary controls (ice cube, chewing a piece of gum, riding a roller coaster, taking a nap, taking a shower, eating a sandwich, waiting in line at a grocery store, taking a dog for a walk), and eight stable controls (a rock, a person's eye color, an ocean, a mountain, an island, a person having toes, a person having a nose, a person having a brain). The order of questions was randomized.

Participants responded to the MAT twice (once during their first visit and again about a week later during their second visit; M = 7.61 days, range = 7 to 16 days). During their first visit, the MAT was the second task (after the temporal reasoning measure). During their second visit, the MAT was the first task. Children and adults were told and tested on their knowledge of the characters' ages, informed about the definitions for each time unit (e.g., by months we mean less

[5]Nine children and four adults completed their first session with only two end-point labels for the certainty measure ("very unsure" to "very sure") and they were not trained how to use the scale to make different kinds of judgments.

than one year), and trained on the certainty scale during both visits before beginning the MAT.

After answering the MAT questions during their first visit, participants completed some exploratory trials aimed at further understanding their reasoning about the duration of mental states. Four "No-Event Emotion" trials assessed whether children and adults would provide differentiated duration estimates for emotions when a person feels an emotion without a precipitating cause (e.g., "Logan feels happy. How long will Logan feel happy?"). Emotions featured were randomized across participants and came from the initial set of 8 emotions. We also tested whether information about fulfillment or blockage would influence desire and preference duration estimates. Participants responded trials where they learned that a person's desire was fulfilled (e.g., "Frances wants strawberries. Frances eats strawberries. How long will Frances want strawberries?"; 2 trials), a person's desire was blocked (e.g., "Quinn wants strawberries. Quinn has no strawberries. How long will Quinn want strawberries?"; 2 trials), a person's preference was fulfilled (e.g., "Ashley likes strawberries. Ashley eats strawberries. How long will Ashley like strawberries?"; 2 trials), and when a person's preference was blocked (e.g., "Andy likes strawberries. Andy has no strawberries. How long will Andy like strawberries?"; 2 trials). During the second visit, participants completed additional exploratory questions after the MAT. They reported how long they expected unspecified emotions ("Logan feels something. How long will Logan feel this?"), desires ("Rory wants something. How long will Rory want this?), and preferences (e.g., Robin likes something. How long will Robin like this?) to last.[6]

Coding and Scoring. We calculated several variables for analyses. That is, within each

[6]During their first session, participants also responded to questions that asked them about whether they thought that desires and preferences also differ in their relation to emotions and decisions. These data will be analyzed in a separate manuscript. During the second session, participants also responded to questions that asked them about social preferences (e.g., "Tanner likes someone (as a friend). How long will Tanner like this person?) and desires (e.g., "Cam wants to be friends with someone. How long will Cam want to be friends with this person?"). These data will be analyzed in a separate manuscript.

session (i.e., Time 1 and Time 2 variables) as well as across sessions (i.e., averaging across Time 1 and Time 2 variables) we created *temporary control variables* (duration estimates, certainty ratings), *emotion variables* (duration estimates, certainty ratings), *desire variables* (duration estimates, certainty ratings), *preference variables* (duration estimates, certainty ratings), *stable control variables* (duration estimates, certainty ratings). All indices were calculated by averaging across the relevant variables for that average. For example, *visit 1 average emotion duration estimate* = (Time 1 startled + Time 1 surprised + Time 1 mad + Time 1 sad + Time 1 worried + Time 1 happy + Time 1 proud + Time 1 ashamed) / 8. When participants failed to answer a question needed to calculate any index described above, we calculated that average out of the number of trials that the participant completed.

A priori we decided to exclude participants who got more than three of the following contrasts incorrect (a rock will last longer than an ice cube; a person's eye color will last longer than chewing a piece of gum; an ocean will last longer than a roller coaster ride; a mountain will last longer than a nap; an island will last longer than a shower; a person will have toes for longer than it takes to eat a sandwich; a person will have a nose for longer than it takes to wait in a line at a grocery store; a person will have a brain for longer than it takes to walk a dog). We excluded one adult participant from analyses for this reason (N = 108).

General Procedure. Participants responded to these measures across two sessions about one week apart. Before the first visit, adult participants and parents of child participant provided written consent. Before both the first and second visits, child participants provided informed assent. Both sessions took place in a quiet room in a university lab. During the first visit, participants completed the temporal reasoning task followed by the MAT. During their second visit, participants completed the MAT. During both sessions (as well as during a third session)

participants completed other social reasoning and individual difference measures. All tasks described in this manuscript were administered via Qualtrics Survey Software (Qualtrics, 2021). To reduce fatigue, an experimenter read the questions aloud to child participants as they read along. Adult participants read the questions to themselves.

Results

Results are presented in four parts. Preliminary analyses tested for age-related differences in temporal reasoning. Next, we examined duration estimates by mental state and by emotion type; we further assessed internal consistency and test-retest reliability. In the third section, we analyze certainty judgments by mental state type, and we report internal consistency and test-retest reliability. In the final section, we conducted exploratory analyses to improve our understanding of children's and adults' beliefs about the duration of emotions, desires, and preferences. Analyses were run in RStudio (Version 3.6.1; R Core Team, 2019).

Temporal Reasoning

Initial analyses examined performance on the temporal reasoning task. We found a significant correlation between temporal reasoning and age (r[106] = .38, 95% confidence interval [CI] [.21, .53], $p < .001$). Still, children's and adults' overall performance on the temporal reasoning task was strong (8- to 10-year-olds: M = .95, 95% CI [.93, .96]; adults: M = .98, 95% CI[.98, .99]). In addition, their performance on each of the subtasks was also good (8- to 10-year-olds: comparisons: M = .99, 95% CI[.98, 1.00]; transformations: M = .85, 95% CI[.80, .90]; ranking: M = .98, 95% CI[.96, 1.00]; Adults: comparisons: M = .99, 95% CI[.98, 1.00]; transformations: M = .97, 95% CI[.95, .99]; ranking: M = .98, 95% CI[.96, 1.00]).[7] A priori, we

[7]For the question, "How many weeks are in a month?" we counted 4, 5, or between the two as correct. When ranking the duration units, four adults provided the accurate rankings but in the opposite order than specified (i.e., they ranked them from longest to shortest instead of shortest to longest). We rescored these responses to be correct.

decided to include temporal reasoning as a covariate in analyses of children's and adults' beliefs about the duration of mental states if it was correlated with age.

Duration Estimates

Mental State Type. We conducted a 2 (age: 8- to 10-year-olds, adults) x 5 (mental state: temporary control, emotion, desire, preference, stable control) repeated measures ANCOVA on duration estimates averaging across Session 1 and Session 2. Age was a between-subjects factor and mental state was a within-subjects factor; a mean-centered temporal reasoning score was included as a covariate. This analysis resulted in a main effect for mental state, $F(4, 420) = 2618.41$, $p < .001$, $\eta_p^2 = .96$ (*Table 2 and Figure 3*). There were no effects for age or temporal reasoning, Fs < 1.13, ps > .290, η_p^2s < .01. To examine the main effect of mental state, we used post-hoc pairwise comparisons (Tukey's HSD corrected). Children and adults rated the temporary controls (e.g., an ice cube) as the most fleeting and the stable controls (e.g., a rock) as the longest lasting (ps < .001). More focally, participants expected preferences (e.g., likes milk) to last longer than emotions (e.g., feels happy) or desires (e.g., wants milk, ps < .001), with emotions and desires judged to last a comparable amount of time ($p = .091$). These patterns also appeared at the individual level, 100% of children and adults judged that preferences would last longer than desires and emotions. In contrast, just over half of participants (58% of children and 57% of adults) expected emotions to be more temporary than desires.

Table 2. Chapter 1: *Study 2 Means [95% Confidence Intervals] of Duration Estimates by Age Group and Item*

	8- to 10-year-olds	Adults	Across Age
Ice Cube: Tatum sees that there is an ice cube on a table.	2.13 [1.84, 2.43]	2.03 [1.96, 2.10]	2.08 [1.93, 2.22]
Gum: Harley is chewing a piece of gum.	2.43 [2.24, 2.63]	2.12 [2.03, 2.20]	2.27 [2.16, 2.38]
Roller Coaster: Sam is on a roller coaster ride.	1.91 [1.82, 2.01]	2.04 [1.84, 2.23]	1.98 [1.87, 2.09]
Napping: Avery just started napping.	2.74 [2.61, 2.87]	2.58 [2.42, 2.74]	2.66 [2.55, 2.76]
Shower: Dakota just started taking a shower.	2.03 [2.00, 2.06]	2.04 [1.99, 2.08]	2.03 [2.01, 2.06]
Sandwich: Austin is eating a sandwich.	1.99 [1.93, 2.05]	2.01 [1.96, 2.06]	2.00 [1.96, 2.04]
Line: Alex is in line to check out at the grocery store.	2.03 [1.95, 2.10]	2.00 [1.93, 2.07]	2.01 [1.96, 2.06]
Dog: Parker is walking the dog.	2.14 [2.05, 2.24]	2.29 [2.13, 2.44]	2.22 [2.13, 2.31]
Temporary Control	2.18 [2.12, 2.24]	2.14 [2.09, 2.18]	2.16 [2.12, 2.19]
Rock: Dylan is on a hike and sees that there is a rock.	8.52 [8.33, 8.71]	8.63 [8.43, 8.84]	8.58 [8.44, 8.72]
Eye Color: Bailey has brown eyes.	8.88 [8.78, 8.97]	8.99 [8.97, 9.01]	8.94 [8.89, 8.98]
Ocean: Sidney is on the coast and sees that there is an ocean.	8.94 [8.89, 9.00]	8.96 [8.92, 8.99]	8.95 [8.92, 8.98]
Mountain: Reagan is on a trip and sees that there is a mountain.	8.85 [8.77, 8.93]	8.73 [8.54, 8.93]	8.78 [8.68, 8.89]
Island: Ray is on an airplane and sees that there is an island.	8.57 [8.36, 8.77]	8.77 [8.62, 8.92]	8.67 [8.55, 8.80]
Toes: Kerry has toes.	8.92 [8.80, 9.04]	9.00 [no variability]	8.96 [8.91, 9.02]
Nose: Piper has a nose.	8.97 [8.94, 9.00]	9.00 [no variability]	8.99 [8.97, 9.00]
Brain: Riley has a brain.	8.97 [8.94, 9.00]	9.00 [no variability]	8.99 [8.97, 9.00]
Stable Control	8.83 [8.77, 8.88]	8.89 [8.83, 8.94]	8.86 [8.82, 8.89]
Startled: Something happened that made Harper feel startled.	1.68 [1.51, 1.85]	1.56 [1.38, 1.74]	1.62 [1.50, 1.74]
Surprised: Something happened that made Corey feel surprised.	1.72 [1.57, 1.87]	1.56 [1.41, 1.70]	1.63 [1.53, 1.74]
Mad: Something happened that made Daryl feel mad.	2.52 [2.36, 2.68]	2.78 [2.60, 2.95]	2.65 [2.53, 2.77]
Sad: Something happened that made Reese feel sad.	2.58 [2.32, 2.83]	3.06 [2.87, 3.25]	2.83 [2.67, 2.99]
Worried: Something happened that made Ellis feel worried.	2.71 [2.52, 2.90]	3.11 [2.90, 3.31]	2.92 [2.77, 3.06]
Happy: Something happened that made Val feel happy.	2.98 [2.72, 3.24]	2.71 [2.54, 2.87]	2.84 [2.68, 2.99]
Proud: Something happened that made Brett feel proud.	3.15 [2.75, 3.56]	3.60 [3.13, 4.07]	3.38 [3.07, 3.69]
Ashamed: Something happened that made Morgan feel ashamed.	2.85 [2.60, 3.10]	3.51 [3.18, 3.84]	3.20 [2.97, 3.41]
Emotion (Feel)	2.52 [2.37, 2.68]	2.73 [2.59, 2.88]	2.63 [2.53, 2.74]
Desire Milk: Rowan wants milk.	2.27 [2.03, 2.51]	2.39 [2.19, 2.58]	2.33 [2.18, 2.48]
Desire Juice: Taylor wants juice.	2.34 [2.14, 2.53]	2.46 [2.25, 2.68]	2.40 [2.26, 2.55]
Desire Pretzels: Ari wants pretzels.	2.66 [2.35, 2.98]	2.55 [2.33, 2.78]	2.61 [2.42, 2.80]
Desire Strawberries: Drew wants strawberries.	2.54 [2.31, 2.77]	2.66 [2.42, 2.91]	2.60 [2.44, 2.77]
Desire to Read: Jules wants to read.	3.00 [2.59, 3.41]	3.04 [2.72, 3.67]	3.02 [2.77, 3.28]
Desire to Dance: Hayden wants to dance.	3.08 [2.61, 3.54]	3.02 [2.71, 3.32]	3.05 [2.78, 3.32]

Desire to Hike: Addison wants to hike.	3.36 [3.10, 3.62]	3.46 [3.19, 3.74]	3.41 [3.22, 3.60]
Desire to Cook: Evan wants to cook.	3.55 [3.04, 4.06]	3.18 [2.81, 3.55]	3.36 [3.05, 3.66]
Desire (Wants)	2.85 [2.62, 3.08]	2.85 [2.66, 3.04]	2.85 [2.70, 2.99]
Preference Milk: Pat likes milk.	7.55 [7.23, 7.87]	7.79 [7.47, 8.12]	7.68 [7.45, 7.90]
Preference Juice: Skyler likes juice.	7.32 [6.99, 7.65]	7.54 [7.16, 7.93]	7.44 [7.18, 7.69]
Preference Pretzels: Jamie likes pretzels.	7.44 [7.10, 7.78]	7.49 [7.06, 7.92]	7.47 [7.20, 7.74]
Preference Strawberries: Carmen likes strawberries.	7.47 [7.18, 7.77]	7.82 [7.48, 8.17]	7.65 [7.42, 7.88]
Preference to Read: Peyton likes to read.	7.62 [7.26, 7.97]	7.47 [6.98, 7.96]	7.54 [7.24, 7.84]
Preference to Dance: Micah likes to dance.	7.34 [6.99, 7.67]	7.27 [6.80, 7.73]	7.30 [7.01, 7.59]
Preference to Hike: Cody likes to hike.	7.18 [6.82, 7.55]	7.04 [6.51, 7.58]	7.11 [6.79, 7.43]
Preference to Cook: Devon likes to cook.	7.63 [7.29, 7.98]	7.54 [7.17, 7.92]	7.59 [7.34, 7.84]
Preference (Likes)	7.44 [7.19, 7.70]	7.50 [7.19, 7.81]	7.47 [7.27, 7.67]

Note. Duration estimates: 1 = seconds, 2 = minutes, 3 = hours, 4 = days, 5 = weeks, 6 = months, 7 = years, 8 = decades, 9 = a person's lifetime or longer. Each item is averaging across Time 1 and Time 2.

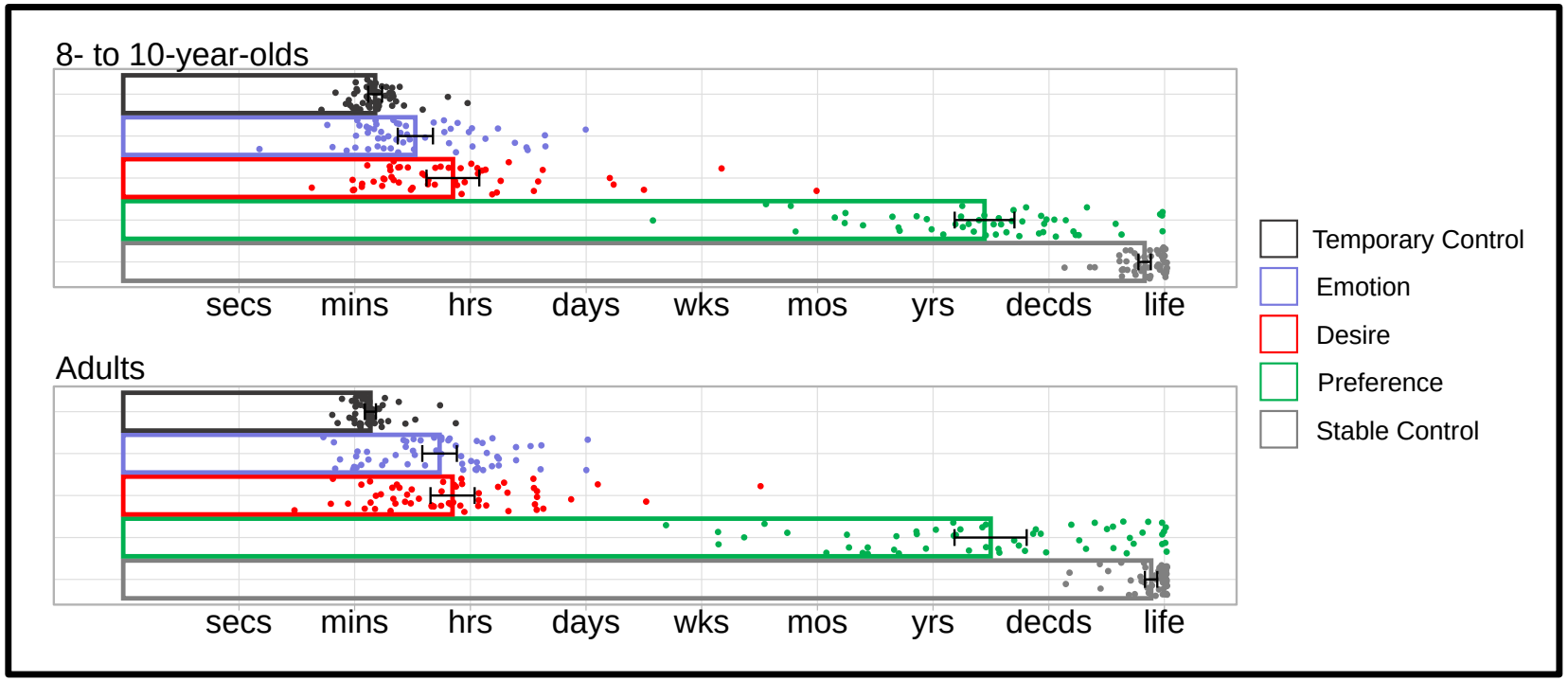

Figure 3. Chapter 1: Study 2 Children's and adults' duration estimates by mental state type. Bar = mean; error bar = 95% confidence interval; dot = jittered individual. Eight- to 10-year-olds and adults differentiated mental states by their time duration: stable control > preferences > desires = emotions > temporary control.

Emotion Type. We conducted a 2 (age) x 8 (emotion: startled, surprised, mad, sad, worried, happy, proud, ashamed) repeated measures ANCOVA on duration estimates averaging across Session 1 and Session 2. Age was a between-subjects factor and mental state was a within-subjects factor; a mean-centered temporal reasoning score was included as a covariate. This analysis resulted in a main effect for emotion, $F(7, 735) = 76.54$, $p < .001$, $\eta_p^2 = .42$, and an Age x Emotion interaction, $F(7, 735) = 3.98$, $p < .001$, $\eta_p^2 = .04$ (*Table 3 and Figure 4*). There were no other significant effects, Fs < 2.56, ps > .113, η_p^2s < .02. Children and adults rated surprised and startled as the most short-lived (ps < .001). Adults also judged that the self-conscious emotions (proud, ashamed) were the longest-lasting (ps < .019; in addition, mad < sad, $p = .048$; happy < worried, sad, ps < .018). In contrast, children's judgements did not follow a consistent pattern (mad, sad < happy, ashamed, proud, ps < .040; worried < proud, $p = .003$).

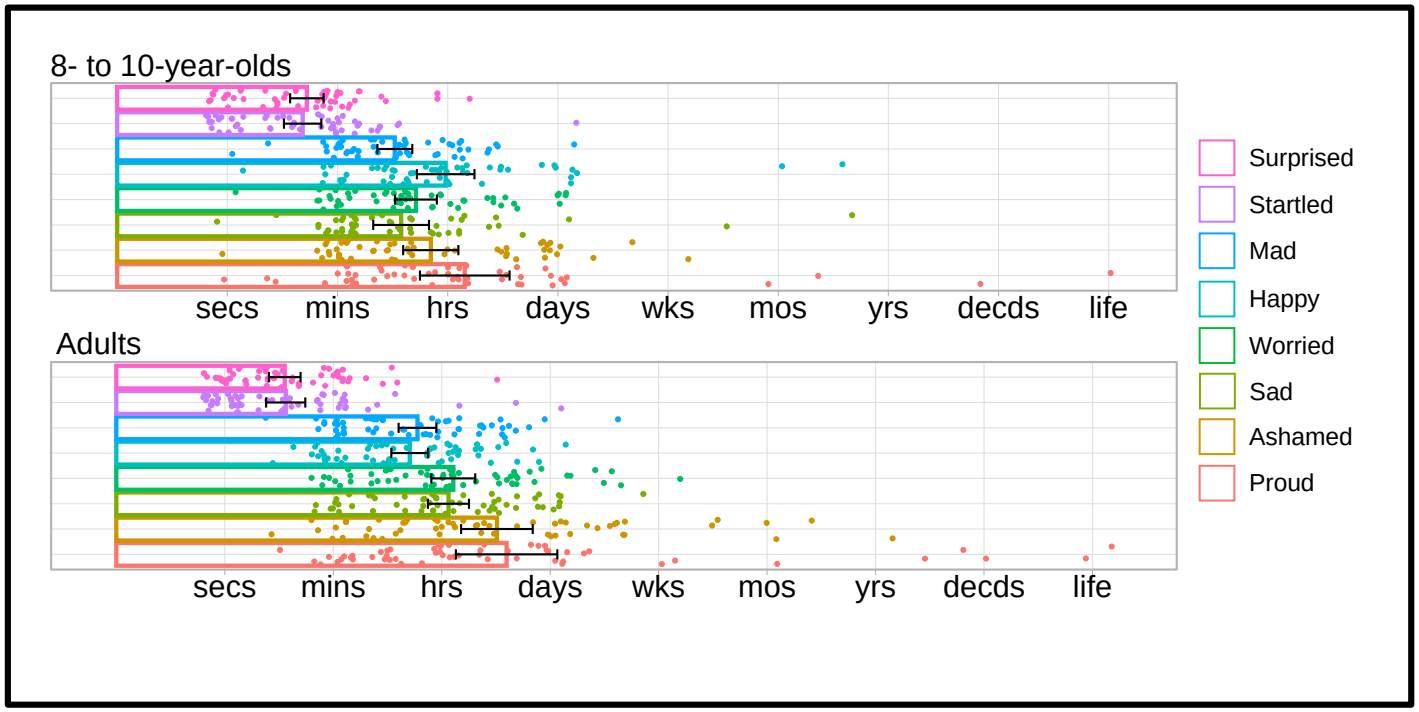

Figure 4. Chapter 1: Study 2 Children's and adults' duration estimates by emotion type. Bar = mean; error bar = 95% confidence interval; dot = jittered individual. Adults differentiated discrete emotions states by their duration: surprised and started were the most fleeting and ashamed and proud were the most stable. Eight- to 10-year-olds also expected surprised and started to be the most fleeting.

Internal Consistency. To test for internal consistency, we examined the Cronbach's alpha at Time 1 and Time 2 for each of the mental states by age group (Time 1: emotion: α = .73, 95% CI[.63, .82]; desire: α = .67, 95% CI [.55, .79]; preference: α = .81, 95% CI[.73, .89]; Time 2: emotion: α = .80, 95% CI [.72, .87]; desire: α = .81, 95% CI[.74, .88]; preference: α = .87 [.81, .92]) and adults (Time 1: emotion: α = .66, 95% CI [.54, .79]; desire: α = .69, 95% CI[.57, .81]; preference: α = .80, 95% CI[.73, .88]; Time 2: emotion: α = .73, 95% CI [.64, .82]; desire: α = .84, 95% CI [.78, .91]; preference: α = .85, 95% CI[.79, .91]). Comparing children's and adults' means and confidence intervals revealed no age-related differences in internal consistency for any mental state.

Test-Retest Reliability. To assess the test-retest reliability of the duration estimates, we conducted correlations between Time 1 and Time 2 judgments separately for each mental state as well as separately for each age group (children: emotion: r[50] = .53, 95% CI [.30, .70], p < .001; desire: r[50] = .70, 95% CI [.53, .82], p < .001; preference: r[50] = .68, 95% CI [.50, .81], p < .001; adults: emotion: r[52] = .76, 95% CI [.62, .85], p < .001; desire: r[52] = .69, 95% CI [.51, .81], p < .001; preference: r[52] = .80, 95% CI [.68, .88], p < .001). Although both age groups exhibited moderate to strong test-retest reliability, comparing means and confidence intervals shows that adults exhibited higher test-retest reliability for emotions than did children.

Between Mental State Duration Correlations. We examined whether participants' duration estimates correlated across mental states separately for each age group (children: emotions vs. desire: r[50] = .07, 95% CI [-.21, .33], p = .642; emotion vs. preference: r[50] = .03, 95% CI [-.24, .30], p = .812; desire vs. preference: r[50] = -.02, 95% CI [-.29, .26], p = .895; adults: emotion vs. desire: r[54] = .43, 95% CI [.18, .62], p = .001; emotion vs. preference: r[54] = -.07, 95% CI [-.32, .20], p = .624; desire vs. preference: r[54] = -.12, 95% CI [-.37, .15], p =

.388). As with Study 1, we found more evidence for within mental state than between mental state relations suggesting that it is not the case that some people just think mental states last a long time and other people believe mental states are short-lived.

Certainty Ratings

Comparisons to the Midpoint. Children and adults exhibited high confidence in their duration estimates (compared to the midpoint of 50): Temporary control (children: $M = 80.86$, $t[51] = 19.27$, $p < .001$, $d = 2.67$; adults: $M = 81.16$, $t[55] = 24.17$, $p < .001$, $d = 3.23$), emotion (children: $M = 72.73$, $t[51] = 12.47$, $p < .001$, $d = 1.73$; adults: $M = 70.84$, $t[56] = 11.96$, $p < .001$, $d = 1.60$), desire (children: $M = 71.91$, $t[51] = 10.77$, $p < .001$, $d = 1.49$; adults: $M = 69.46$, $t[56] = 11.35$, $p < .001$, $d = 1.52$), preference (children: $M = 76.16$, $t[51] = 13.47$, $p < .001$, $d = 1.87$; adults: $M = 75.85$, $t[56] = 14.95$, $p < .001$, $d = 2.00$), stable control (children: $M = 92.12$, $t[51] = 36.52$, $p < .001$, $d = 5.06$; adults: $M = 93.73$, $t[56] = 67.37$, $p < .001$, $d = 9.00$).

Mental State Type. To test for differences by age and mental state type, we conducted an exploratory 2 (age) x 5 (mental state) repeated measures ANOVA on certainty ratings averaging across Time 1 and Time 2. Participant age was a between-subjects factor and mental state was a within-subjects factor. This analysis yielded a main effect for mental state, $F(4, 424) = 202.82$, $p < .001$, $\eta_p^2 = .66$ (*Table 4*). There were no other significant effects, Fs < 1.67, ps > .155, η_p^2s < .02. Post-hoc pairwise comparisons (Tukey's HSD corrected) showed that participants exhibited higher confidence about stable controls > temporary controls > preferences > emotions, desires (ps < .001; emotions = desires, $p = .735$).

Table 3. Chapter 1: *Study 2 Means [95% Confidence Intervals] of Certainty Ratings by Age Group and Item*

	8- to 10-year-olds	Adults	Across Age
Ice Cube: Tatum sees that there is an ice cube on a table.	81.51 [77.27, 85.75]	84.04 [80.46, 87.63]	82.82 [80.10, 85.55]
Gum: Harley is chewing a piece of gum.	77.46 [72.80, 82.12]	77.42 [73.99, 80.85]	77.44 [74.63, 80.25]
Roller Coaster: Sam is on a roller coaster ride.	83.06 [78.74, 87.38]	85.22 [81.68, 88.76]	84.18 [81.44, 86.92]
Napping: Avery just started napping.	78.10 [74.06, 82.13]	74.05 [69.96, 78.15]	76.00 [73.14, 78.86]
Shower: Dakota just started taking a shower.	83.48 [79.59, 87.37]	84.70 [81.50, 87.90]	84.11 [81.65, 86.57]
Sandwich: Austin is eating a sandwich.	83.56 [78.97, 88.14]	83.45 [80.27, 86.62]	83.50 [80.79, 86.21]
Line: Alex is in line to check out at the grocery store.	79.36 [74.74, 83.98]	83.12 [79.43, 86.80]	81.31 [78.40, 84.21]
Dog: Parker is walking the dog.	80.35 [76.12, 84.58]	77.24 [73.44, 81.04]	78.74 [75.93, 81.54]
Temporary Control	80.86 [77.64, 84.07]	81.16 [78.57, 83.74]	81.01 [79.00, 83.03]
Rock: Dylan is on a hike and sees that there is a rock.	87.14 [83.42, 90.87]	90.12 [86.83, 93.40]	88.69 [86.24, 91.13]
Eye Color: Bailey has brown eyes.	91.60 [88.21, 94.99]	95.48 [94.16, 96.80]	93.61 [91.83, 95.39]
Ocean: Sidney is on the coast and sees that there is an ocean.	93.38 [90.56, 96.21]	94.40 [92.40, 96.40]	93.91 [92.23, 95.59]
Mountain: Reagan is on a trip and sees that there is a mountain.	91.53 [87.89, 95.17]	93.77 [91.67, 95.86]	92.69 [90.65, 94.73]
Island: Ray is on an airplane and sees that there is an island.	86.51 [82.28, 90.74]	92.11 [89.95, 94.26]	89.41 [87.07, 91.76]
Toes: Kerry has toes.	94.61 [92.04, 97.17]	93.13 [91.37, 94.88]	93.84 [92.32, 95.35]
Nose: Piper has a nose.	95.96 [93.55, 98.37]	94.81 [93.59, 96.03]	95.37 [94.06, 96.67]
Brain: Riley has a brain.	96.25 [94.41, 98.09]	95.99 [94.96, 97.02]	96.12 [95.10, 97.13]
Stable Control	92.12 [89.81, 94.44]	93.73 [92.42, 95.03]	92.95 [91.66, 94.24]
Startled: Something happened that made Harper feel startled.	77.32 [72.99, 81.65]	75.99 [72.11, 79.87]	76.63 [73.78, 79.48]
Surprised: Something happened that made Corey feel surprised.	77.10 [72.95, 81.24]	78.25 [74.74, 81.76]	77.69 [75.04, 80.35]
Mad: Something happened that made Daryl feel mad.	71.14 [65.33, 76.96]	68.29 [64.07, 72.50]	69.66 [66.16, 73.17]
Sad: Something happened that made Reese feel sad.	70.46 [66.12, 74.80]	66.39 [61.85, 70.94]	68.35 [65.23, 71.47]
Worried: Something happened that made Ellis feel worried.	69.28 [64.12, 74.43]	65.86 [61.73, 69.98]	67.50 [64.27, 70.74]
Happy: Something happened that made Val feel happy.	73.23 [68.93, 77.48]	72.38 [68.03, 76.72]	72.79 [69.79, 75.78]
Proud: Something happened that made Brett feel proud.	72.96 [68.22, 77.70]	72.21 [67.89, 76.52]	72.57 [69.43, 75.71]
Ashamed: Something happened that made Morgan feel ashamed.	70.38 [65.95, 74.80]	67.38 [63.03, 71.74]	68.82 [65.75, 71.89]
Emotion (Feel)	72.73 [69.07, 76.39]	70.84 [67.35, 74.33]	71.75 [69.26, 74.24]
Desire Milk: Rowan wants milk.	72.49 [67.92, 77.06]	68.63 [64.18, 73.07]	70.49 [67.33, 73.64]
Desire Juice: Taylor wants juice.	74.60 [69.94, 79.25]	70.05 [66.16, 73.95]	72.24 [69.24, 75.24]
Desire Pretzels: Ari wants pretzels.	72.34 [67.72, 76.95]	68.01 [63.75, 72.26]	70.09 [66.98, 73.20]
Desire Strawberries: Drew wants strawberries.	70.55 [65.16, 75.94]	69.50 [65.75, 73.25]	70.00 [66.81, 73.20]
Desire to Read: Jules wants to read.	70.74 [65.71, 75.77]	69.33 [65.52, 73.14]	70.01 [66.93, 73.09]
Desire to Dance: Hayden wants to dance.	73.23 [68.62, 77.84]	70.38 [66.12, 74.63]	71.75 [68.66, 74.84]

Desire to Hike: Addison wants to hike.	70.96 [65.85, 76.08]	71.04 [66.67, 75.40]	71.00 [67.71, 74.29]
Desire to Cook: Evan wants to cook.	70.35 [65.52, 75.17]	68.79 [64.61, 72.96]	69.54 [66.41, 72.66]
Desire (Wants)	71.91 [67.82, 75.99]	69.46 [66.03, 72.90]	70.64 [68.02, 73.26]
Preference Milk: Pat likes milk.	74.59 [69.49, 79.69]	75.15 [69.77, 80.54]	74.77 [71.38, 78.16]
Preference Juice: Skyler likes juice.	75.15 [69.77, 80.54]	74.41 [70.03, 78.79]	74.77 [71.38, 78.16]
Preference Pretzels: Jamie likes pretzels.	73.91 [69.50, 78.32]	72.63 [68.03, 77.22]	73.25 [70.10, 76.39]
Preference Strawberries: Carmen likes strawberries.	76.23 [70.75, 81.72]	76.70 [72.39, 90.00]	76.47 [73.07, 79.87]
Preference to Read: Peyton likes to read.	80.32 [76.22, 84.42]	79.79 [76.31, 83.26]	80.04 [77.41, 82.67]
Preference to Dance: Micah likes to dance.	76.72 [72.40, 81.04]	76.69 [72.76, 80.61]	76.70 [73.84, 79.57]
Preference to Hike: Cody likes to hike.	76.08 [71.33, 80.83]	76.68 [72.55, 80.81]	76.39 [73.31, 79.47]
Preference to Cook: Devon likes to cook.	76.32 [72.00, 80.63]	76.99 [73.08, 80.91]	76.67 [73.81, 79.52]
Preference (Likes)	76.16 [72.27, 80.06]	75.85 [72.39, 79.32]	76.00 [73.45, 78.56]

Note. Certainty ratings: 0 = Very unsure to Very sure. Each item is averaging across Time 1 and Time 2.

Internal Consistency. To test for internal consistency of participants' certainty ratings, we examined the Cronbach's alpha at Time 1 and Time 2 for each of the mental states (Time 1: emotion: α = .83, 95% CI[.76, .90]; desire: α = .87, 95% CI [.81, .92]; preference: α = .85, 95% CI[.78, .91]; Time 2: emotion: α = .87, 95% CI [.81, .92]; desire: α = .92, 95% CI[.89, .96]; preference: α = .92 [.89, .95]) and adults (Time 1: emotion: α = .86, 95% CI [.81, .92]; desire: α = .85, 95% CI[.79, .91]; preference: α = .89, 95% CI[.85, .94]; Time 2: emotion: α = .94, 95% CI [.91, .96]; desire: α = .94, 95% CI [.92, .96]; preference: α = .92, 95% CI[.89, .95]). Comparing means and confidence intervals revealed that adults provided more internally consistent certainty ratings for emotions during Time 2 than did children.

Test-Retest Reliability. To assess test-retest reliability of the certainty ratings, we conducted correlations between Time 1 and Time 2 certainty ratings separately for each mental state as well as separately for each age group (children: emotion: $r[50] = .72$, 95% CI [.56, .83], $p < .001$; desire: $r[50] = .73$, 95% CI [.57, .83], $p < .001$; preference: $r[50] = .73$, 95% CI [.58, .84], $p < .001$; adults: emotion: $r[53] = .76$, 95% CI [.62, .86], $p < .001$; desire: $r[53] = .68$, 95% CI [.50, .80], $p < .001$; preference: $r[53] = .74$, 95% CI [.59, .84], $p < .001$). Thus, both children and adults exhibited good test-retest reliability on their certainty ratings.

Between Mental State Certainty Correlations. We examined correlations between participants' certainty ratings across mental states (children: emotions vs. desire: $r[50] = .93$, 95% CI [.89, .96], $p < .001$; emotion vs. preference: $r[50] = .81$, 95% CI [.69, .89], $p < .001$; desire vs. preference: $r[50] = .87$, 95% CI [.78, .92], $p < .001$; adults: emotions vs. desire: $r[54] = .95$, 95% CI [.91, .97], $p < .001$; emotion vs. preference: $r[54] = .78$, 95% CI [.64, .86], $p < .001$; desire vs. preference: $r[54] = .82$, 95% CI [.70, .89], $p < .001$). In other words, individuals were consistently very certain when providing mental state duration estimates.

Exploratory Analyses

Desire and Preference: Fulfillment versus Blockage. We ran a 2 (age) x 2 (mental state: desire, preference) x 3 (ending: standard, fulfilled, blocked) repeated measures ANCOVA on duration estimates. Age was a between-subjects factor; mental state and ending were within-subjects factors; a mean-centered temporal reasoning score was included as a covariate. By "standard" ending we mean the core desire and preference trials included in the MAT where no information was provided as to whether the mental state was fulfilled or blocked. Due to a programming error, 4 participants did not have enough data to be included in this analysis (2 8- to 10-year-olds, 2 adults $N = 106$). This analysis resulted in a main effect for mental state, $F(1, 101) = 491.40$, $p < .001$, $\eta_p^2 = .83$, and ending, $F(2, 202) = 5.37$, $p = .005$, $\eta_p^2 = .04$, qualified by a Mental state x Ending interaction, $F(2, 202) = 27.71$, $p < .001$, $\eta_p^2 = .22$.[8] Of key interest was whether participants' expectations that preferences last longer than desires resulted from an assumption that desires will be fulfilled quickly. We did not find evidence for this alternative explanation. For all ending types (standard, fulfilled, blocked), children and adults rated preferences to last longer than desires, *p*s < .001 (see *Table 4*). Still, within mental state, ending type did shape participants' duration estimates. That is, children and adults expected blocked desires to last longer than fulfilled desires or the standard desire trials (*p*s < .001; fulfilled = standard, $p = .722$). As well, children and adults expected preferences in standard trials to last longer than preferences that were blocked or fulfilled (*p*s < .001; blocked = fulfilled, $p = .170$).

[8]There was also a main effect for temporal reasoning, $F(1, 101) = 4.41$, $p = .038$, $\eta_p^2 = .04$, and a Temporal reasoning x Ending interaction, $F(2, 202) = 3.60$, $p = .029$, $\eta_p^2 = .03$. When averaging across mental states (desire, preference), participants who had better temporal reasoning scores expected blocked desires and preferences to last longer than those with weaker temporal reasoning skills ($r[106] = .25$, 95% CI [.07, .42], $p = .009$). Temporal reasoning was unrelated to participants' beliefs about the duration of fulfilled desires and preferences ($r[106] = .15$, 95% CI [-.04, .33], $p = .114$) or standard desires and preferences ($r[106] = -.05$, 95% CI [-.24, .14], $p = .590$).

Table 4. Chapter 1: *Study 2 Means [95% Confidence Intervals] of Exploratory Variables Duration Estimates by Age Group and Item*

	8- to 10-year-olds	Adults	Across Age
Emotion No Event (Feels)	2.35 [2.20, 2.50]	2.87 [2.64, 3.10]	2.62 [2.47, 2.77]
Desire Fulfilled (Wants)	2.92 [2.38, 3.47]	2.70 [2.30, 3.10]	2.81 [2.48, 3.14]
Preference Fulfilled (Likes)	6.30 [5.69, 6.92]	6.78 [6.13, 7.44]	6.55 [6.11, 7.00]
Desire Blocked (Wants)	3.68 [3.29, 4.07]	3.82 [3.31, 4.33]	3.75 [3.43, 4.08]
Preference Blocked (Likes)	5.89 [5.37, 6.42]	6.54 [5.94, 7.15]	6.23 [5.83, 6.63]
Unspecified Emotion (Feels something)	2.56 [2.20, 2.92]	2.54 [2.24, 2.83]	2.55 [2.32, 2.77]
Unspecified Desire (Wants Something)	3.72 [3.18, 4.24]	3.19 [2.80, 3.58]	3.44 [3.12, 3.77]
Unspecified Preference (Likes Something)	6.96 [6.56, 7.36]	6.39 [5.77, 7.01]	6.67 [6.30, 7.04]

Note.

Duration estimates: 1 = seconds, 2 = minutes, 3 = hours, 4 = days, 5 = weeks, 6 = months, 7 = years, 8 = decades, 9 = a person's lifetime or longer.

Emotion: Event versus No Event. We conducted a 2 (age) x 2 (type: event, no event) repeated measures ANCOVA on duration ratings for emotions. Age was a between-subjects factor and type was a within-subjects factor; a mean-centered temporal reasoning score was included as a covariate. This analysis resulted in a main effect for age, $F(1, 105) = 6.45$, $p = .013$, $\eta_p^2 = .06$, and type, $F(1, 105) = 9.30$, $p = .003$, $\eta_p^2 = .08$, qualified by an Age x Type interaction, $F(1, 105) = 4.48$, $p = .037$, $\eta_p^2 = .04$. Eight- to 10-year-olds judged that emotions with a precipitating event would last longer than emotions without a cause ($p < .001$), but adults did not make this distinction ($p = .588$). In addition, adults expected emotions with no causal event to last longer than did children ($p = .002$; see *Table 4*).

No-Event Emotion versus Standard Desire and Preference Trials. We conducted a 2 (age) x 3 (mental state: no-event emotion, desire, preference) repeated measures ANCOVA on duration estimates, with age as a between-subjects factor, mental state as a within-subjects factor, and mean-centered temporal reasoning score as a covariate. This resulted in a main effect for age, $F(1, 105) = 4.15$, $p = .044$, $\eta_p^2 = .04$, and mental state, $F(2, 210) = 970.59$, $p < .001$, $\eta_p^2 = .90$. The Age x Mental state interaction was not significant, $F(2, 210) = 1.71$, $p = .183$, $\eta_p^2 = .02$. Adults expected mental states to last longer than did children, $p = .044$. Consistent with primary analyses, participants expected preferences to last longer than desires and emotions (here, with no precipitating event), ps $< .001$. In addition, children and adults expected emotions (with no precipitating event) to last a comparable amount of time as desires, $p = .158$ (see *Table 4*).

Unspecified Mental States. We conducted a 2 (age) x 3 (mental state: emotion, desire, preference) repeated measures ANOVA on duration estimates for unspecified trials (i.e., Sam feels something; Alex wants something; Joe likes something). Two adults who did not complete these trials ($N = 106$). This analysis resulted in a main effect for mental state, $F(206) = 207.11$, p

< .001, η_p^2 = .67. All other effects were null, *F*s < 3.42, *p* > .067, η_p^2s < .03. When provided with no information beyond mental state term, participants judged that preferences would last longer than desires, and that desires would last longer than emotions (*p*s < .001; see *Table 4*).

Discussion

Study 2 provided a replication and extension of Study 1. Consistent with Study 1, adults believed that emotions and desires would be shorter lived than preferences. Adding to Study 1, 8- to 10-year-olds provided comparable judgments to adults (although, unlike adults, children did not expect self-conscious emotions to be comparatively long-lasting). We further documented that children and adults not only reported confidence in their estimates, but their judgments were internally consistent and reliable over time. The inclusion of additional exploratory questions added important clarifications. That is, children's and adults' duration rank order (emotions, desires < preference), remained constant when emotion trials specified no precipitating event, when stated desires and preferences were blocked or fulfilled, or when trials included only the mental state term with an unspecified referent (e.g., Casey wants something.).

General Discussion

The current research demonstrates that children and adults have theory-like intuitions about the duration of mental states. Across two studies, adults judged that emotions and desires are more temporary than preferences (with weaker evidence that they may also view emotions as more temporary than desires). By 8 to 10 years of age, children held similar, but less nuanced beliefs. That is, whereas children's judgments about the duration of the overarching mental state categories matched that of adults (emotions, desires < preferences), they did not expect self-conscious emotions (i.e., ashamed, proud) to be consistently longer-lasting than the other emotions (i.e., happy, sad, mad, worried, surprised, and startled). Importantly, although there was

clear individual variability (e.g., some participants thought that preferences would last years but others thought that they would last a lifetime), the rank orders were robust at the level of individual participant and item. Moreover, both children and adults exhibited high confidence in their judgments and provided internally consistent and reliable responses across a 1-week delay.

Beliefs about the Duration of Mental States

Research on developmental changes in mental state reasoning has primarily focused on documenting at what age children first begin to appreciate that a person's mental states (e.g., beliefs, desires, emotions, intentions, preferences) drive their actions (e.g., search behaviors, facial expressions; Wellman, 2014). This emphasis has led researchers to know a great deal about the developmental trajectory and mechanisms of change in children's and adults' reasoning about links between and among mind, behavior, and the world (Milligan et al., 2007; Devine & Hughes, 2014; Wellman et al., 2001). Unfortunately, however, this concentration on early knowledge about mental states has also been at the detriment of exploring other more complex, yet everyday, aspects of children's and adults' mental state reasoning (Kramer & Lagattuta, in press; Lagattuta & Kramer, 2021; Lagattuta et al., 2015).

Here, we add to the growing literature on more advanced aspects of theory of mind by taking seriously the notion that mental states do not always occur in isolation, but rather unfold over time. In contrast to prior work showing that 8- to 10-year-olds and adults hold different beliefs about the interplay between mental states and time (Lagattuta & Kramer, 2021; Lara et al., 2019; Guttentag & Ferrell, 2008; McCormack & Feeney, 2015), we largely found continuity in their reasoning about the time duration of emotions, desires, and preferences. These findings are particularly intriguing because children younger than age 8 do not have complete mastery of temporal units (e.g., they might think that 7 hours is shorter than 9 minutes; Tillman & Barner,

2015). That is, our data show that children can map mental states onto their time duration (in the same way as adults do) when they are first solidifying their knowledge about temporal terms. We are currently developing a modified scale to use with younger children (e.g., 4- to 7-year-olds) that uses pictures and verbal labels (e.g., a really short time, a really long time) as opposed to specific temporal units (e.g., seconds, weeks, decades). This can elucidate whether even younger children preserve these rank order distinctions, potentially indicating that knowledge about the relative time course of different mental states is also evident in younger children's minds.

Although we did not document large age differences, we did discover that whereas adults viewed some discrete emotions as having unique time durations, children did not consistently share these same beliefs. More specifically, children and adults judged that surprised and startled would be the most short-lived emotions, but only adults also viewed self-conscious emotions (shame and pride) as particularly long lasting. These findings fit with the more general emotion understanding literature showing that children's knowledge of self-conscious emotions is protracted (Lagattuta & Thompson, 2007). For example, children have more difficulty labeling shame and pride emotional expressions than they do more basic emotions (e.g., happy, sad; e.g., Nelson & Russel, 2012; but see Tracy et al., 2005 for evidence that children can discriminate happy and proud). In addition, they struggle to understand the situations that lead people to experience pride, shame, and embarrassment (Banerjee, 2002; Ferguson et al., 1991; Thompson, 1987). Here, we add to this literature by showing that children also exhibit less awareness than adults about how emotions that are reflective of the self can endure over time.

Distinctions Among Mental States

In some ways, children's and adults' reliable beliefs that desires (wants) are shorter-lived than preferences (likes) are surprising because the two mental states share common features.

They both manifest as affective, appetitive stances (I *want* candy versus I *like* candy) and children develop general understanding of them early in development. It is also common in the scientific literature and in everyday conversations to use words that signal desires (e.g., "want") and those that denote preferences (e.g., "like") interchangeably. Despite these commonalities, the neurobehavioral addiction literature suggests that wanting and liking are distinct processes (Berridge & Robinson, 2016): Dopamine fluctuates with increases in wanting, but not liking (Berridge et al., 1989). Our findings add to this literature on like-want distinctions by illustrating that children and adults hold dissimilar temporal representations of them. Importantly, in Study 2, we found that children and adults did not distinguish desires from preferences simply because they assumed that desires would be fulfilled. Indeed, children and adults consistently judged that preferences would last longer than desires when no fulfillment information was provided, when the desire and the preference were fulfilled, and when the desire and the preference were blocked. It will be interesting in future work to examine how children and adults conceptualize other differences between wants and likes (e.g., Is liking or wanting more linked to behaviors?).

We can imagine circumstances under which beliefs about the time course of desires and preferences may get distorted and that these confusions can shape in-the-moment decisions. In our work, participants thought that wanting something would dissipate more quickly than liking it, but other studies find that people rely heavily on their current desires (e.g., feeling hungry or thirsty) when predicting their future states (suggesting a belief that desires can be long-lasting). This finding has been interpreted as people's difficulty disengaging from their current experiences to think about a future state (Atance & Meltzoff, 2006; Gilbert et al., 2002; Kramer et al., 2017; Loewenstein et al., 2003; Mahy et al., 2014; Nisbett & Kanouse, 1968; Read & van Leeuwen, 1998). Perhaps another component of this struggle is that children and adults have

trouble separating whether their current state is what they want right now or what they will like for a long time (i.e., difficulties with mapping their theories of the duration of mental states onto their actual experiences). Indeed, it is worth noting that in our study we took a third-person approach, but ideas about the time course of mental states may differ in a first-person context. For example, children and adults have an easier time thinking about future desires and preferences when they are considering another person's experience rather than their own (Bauckham et al., 2019; Belanger et al., 2014; Mahy et al., 2020; Renoult et al., 2016). The current research suggests that one intervention for better decision making could be as simple as reminding people of their own general theories (e.g., desires subside, even when unfulfilled).

The current research provides weak evidence that children and adults represent emotions and desires as having distinct time durations. In Study 1, adults viewed desires as longer-lasting than emotions. In Study 2, although this same pattern held numerically for both children and adults, the difference was not significant. Moreover, in Study 1 adults' emotion and desire judgments were unrelated as were children's in Study 2 (but, in Study 2 the adults' emotion and desire duration estimates correlated). Thus, to better understand whether children and adults view emotions and desires as differing in their time course may require larger samples to detect smaller effects than those we were able to observe here. Still, the lack of distinction may also be explained by the discrete emotions that we included (e.g., that self-conscious emotions are longer lasting than desires but other emotions are not). Indeed, when we analyzed participants' beliefs about the duration of emotions and desires without the discrete emotion or desire referent (i.e., the unspecified emotion and unspecified desire trials), participants rated emotions as more temporary than desires. It is also possible that duration is not a salient difference between emotions and desires because they differ in several other aspects (e.g., their causes,

consequences, and physiological markers), making time course less criterial for separating them.

Limitations and Future Directions

We created a paradigm to test a new question; thus, it is essential to consider potential constraints on the generality of these findings (Simons et al., 2017). We expect these patterns to replicate in other undergraduate and 8- to 10-year-olds from WEIRD populations (western, educated, industrialized, rich, democratic; Henrich et al., 2010). It is unknown whether these results would replicate in other age groups or in non-WEIRD samples. The culture in which individuals grow up likely shapes their thinking about the interplay between mental states and time. The extent to which emotions are shared and displayed may influence beliefs about their duration; the centrality of preferences to the self could change judgments of stability; and the value placed on self-control could affect theories about the time course of desires. We encourage work testing similarities and differences in beliefs about the time course of mental states across cultures, especially from a developmental perspective. We should also interrogate how our results may have changed had we chosen different items to represent our dependent variables (e.g., Sam likes *math* versus Sam likes *pretzels*). Importantly, we do not believe that our conclusions are dependent on these decisions. Although there were differences in participants' beliefs about duration of certain items (e.g., wanting milk is shorter-lived than wanting to hike), the general rank order (e.g., preferences last longer than desires) remained regardless of referent. It will still be important to vary the object of individuals' preferences and desires further to gain better understanding about lay beliefs regarding how long different mental states last.

We are excited about the many directions that research on children's and adults' beliefs about the duration of mental states could take. For example, it would be informative to manipulate aspects of the mental state, such as its intensity (e.g. Sam wants pretzels a lot versus a

little bit), origin recency (e.g., Sam just started to like milk versus he has liked milk for a long time), and cause (e.g., Sam feels sad because his dog is lost versus because he spilled milk). We narrowed our focus to emotions, desires, and preferences (see Table A.1 for descriptive data about the time duration of some additional mental states and characteristics), but expanding systematic inquiry to other mental states and referents would be important. For example, ideas about the time course of varying kinds of beliefs (e.g., political, religious, scientific) could enhance scientific understanding about growing ideological divides. Modifying characteristics of the mental-state experiencer, such as their age, social group membership, or intersectional identities—and assessing judgments from both first-person and third-person perspectives—could provide insight into whether children and adults believe that the time course of mental states depends upon the agent, including whether they expect outgroup differences.

We introduced the concept of beliefs about the duration of mental states by suggesting that such ideas may matter for wellbeing and decision making. For example, believing that negative emotions (e.g., sadness) are particularly long-lasting may contribute to the creation and maintenance of mental health issues (e.g., depression; see also Ford & Gross, 2019). Anticipating that desires last a long time may exacerbate addiction or dampen the ability to delay gratification. Moreover, the assumption that preferences are stable could influence how people make decisions with long-term consequences (e.g., buying a house). Importantly, although duration rank orders were consistent across participants, there was clear variability in these judgments. It will be intriguing to test whether these individual differences meaningfully predict certain outcomes. Still, it is possible that time course judgments participants provide when not currently experiencing a salient state are less relevant for overall wellbeing and decision making than beliefs assessed during a more emotionally-volatile time (e.g., beliefs about the duration of

sadness may vary depending upon whether the person is currently feeling sad). Although we documented consistency in judgments across a 1-week delay, participants were presumably in non-arousing situations at both time points. Thus, it will be necessary to test relations between beliefs and various outcomes across multiple contexts.

Conclusions

From moment to moment, mental states can fluctuate. For example, a person can feel happy thinking about a past joyous event, but feel sad upon realizing that the event has ended. Desires shift as well: A child can want a cookie, but then decide that she actually does not want it soon after the desire began. In contrast, other mental experiences, may feel relatively stable. Someone who likes chocolate will probably continue to like it for a long time. We discovered that by the age of 8 to 10 years, children and adults appreciate these temporal components of mental states. Said another way, individuals judge that mental state terms (feeling, wanting, liking) carry temporal duration information. Children and adults viewed preferences (liking) as substantially longer lasting than emotions (feeling) and desires (wanting), with some evidence that they think about the time course of emotions and desires in distinct ways as well. Taken together, these findings highlight the utility and necessity of considering how children and adults incorporate temporal features into their mental state reasoning.

References

Amir, O., & Levav, J. (2008). Choice construction versus preference construction: The instability of preferences learned in context. *Journal of Marketing Research, 45*, 145-158. doi: 10.1509/jmkr.45.2.145

Arsenio, W. F., Gold, J., & Adams, E. (2006). Children's conceptions and displays of moral emotions. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 581-610). Psychology Press.

Atance, C.M, Bélanger, M., & Meltzoff, A.N. (2010). Preschoolers' understanding of others' desires: Fulfilling mine enhances my understanding of yours. *Developmental Psychology, 46* 1505-1513. doi: 10.1037/a0020374

Atance, C.M., & Meltzoff, A.N. (2006). Preschoolers' current desires warp their choices for the future. *Psychological Science, 17*, 583-587. doi: 10.1111/j.1467-9820.2006.01748.x

Banerjee, R. (2002). Audience effects on self-presentation in childhood. *Social Development, 11*, 487-507. doi: 10.1111/1467-9507.00212

Barden, R.C., Zelko, F.A., Duncan, S.W., & Masters, J.C. (1980). Children's consensual knowledge about the experiential determinants of emotions. *Journal of Personality and Social Psychology, 39*, 968-976.

Bauckham, G., Lambert, R., Atance, C., Davidson, P.S., Taler, V., & Renoult, L. (2018). Predicting our own and others' future preferences: The role of social distance. *Quarterly Journal of Experimental Physiology, 72*, 634-642. doi: 10.1177/1747021818763573

Belanger, M.J., Atance, C.M., Varghese, A.L., Nguyen, V., & Vendetti, C. (2014). What will I like best when I'm all grown up? Preschoolers' understanding of future preferences. *Child Development, 85*, 2419-2431. doi: 10.1111/cdev.12282

Bernstein, D.M. (2018). Lifespan theory of mind: A call for broader perspectives and more integration. *Zeitschrift für Psychologie, 226*, 85-86. doi: 10.1027/2151-2604/a000324

Berridge, K.C. & Robinson, T.E. (2016). Liking, wanting and the incentive-sensitization theory of addiction. *American Psychologist, 71*, 670-679. doi: 10.1037/amp0000059

Berridge, K.C., Venier, I.L., & Robinson, T.E. (1989). Taste reactivity analysis of 6-hydroxydopamine-induced aphagia: Implication for anhedonia hypotheses of dopamine function. *Behavioral Neuroscience, 103*, 36-45. doi: 10.1037/0735-7044.103.1.36

Brandone, A. C., & Klimek, B. (2018). The developing theory of mental state control: Changes in beliefs about the controllability of emotional experience from elementary school through adulthood. *Journal of Cognition and Development, 19*, 509-531. doi: 10.1080/15248372.2018.1520711

Cassidy, K., Cosetti, M., Jones, R., Kelton, E., Rafal, V.M., Richman, L.,…Stanhaus, H. (2005). Preschool children's understanding of conflicting desires. *Journal of Cognition and Development, 6*, 427-454. doi: 10.1207/s15327647jcd0603_6

Choi, Y.J., Mou, Y., Luo, Y. (2018). How do 3-month infants attribute preference to human agent? *Journal of Experimental Child Psychology, 172*, 96-106. Do: 10.1016/j.jecp.2018.03.004

Devine, R.T., & Hughes, C. (2014). Relations between false belief and understanding and executive function in early childhood: A meta-analysis. *Child Development, 85*, 1777-1794. doi: 10.1111/cdev.12237

Diesendruck, G., Salzer, S., Kushnir, T., & Xu, F. (2015). When choices are not personal: The effect of statistical and social cues on children's inferences about the scope of preference. *Journal of Cognition and Development, 16*, 370-380. doi:

10.1080/15248372.2013848870

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Ferguson, T. J., Stegge, H., & Damhuis, I. (1991). Children's understanding of guilt and shame. *Child Development, 62*, 827-839. doi: 10.111/j.1467-8624.1991.tb01572.x

Ford, B. Q., & Gross, J. J. (2019). Why beliefs about emotion matter: An emotion-regulation perspective. *Current Directions in Psychological Science, 28,* 74-81. doi: 10.1177/0963721418806697

Frijda, N. H., Mesquita, B., Sonnemans, J., & van Goozen, S. H. M. (1991). The duration of affective phenomena or emotions, sentiments and passions. *International Review of Studies on Emotion, 1*. Wiley.

Gilbert, D.T., Gill, M.J., & Wilson, T.D. (2002). The future is now: Temporal correction in affective forecasting. *Organization Behavior and Human Decision Processes, 88*, 430-444. doi: 10.1006/obhd.2001.2982

Goldfarb, D., Lagattuta, K.H., Kramer, H.J., Kennedy, K., & Tashjian, S.M. (2017). When your kind cannot live here: How generic language and criminal sanctions shape social categorization. *Psychological Science, 28*, 1597-1609. doi: 10.1177/0956797617714827

Guttentag, R., & Ferrell, J. (2008). Children's understanding of anticipatory regret and disappointment. *Cognition and Emotion, 22*, 815-832. doi: 10.1080/02699930701541542

Gweon, H., Tenenbaum, J.B., & Schulz, L.E. (2010). Infants consider both the sample and

the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 9066-9071. doi: 10.1073/pnas.1003095107

Harris, P.L. (2010). Children's understanding of emotion. In M. Lewis, J. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions* (pp. 320-331). The Guilford Press.

Harris, P.L., Guz, G.R., Lipian, M.S., & Man-Shu, Z. (1985). Insight into the time course of emotion among Western and Chinese children. *Child Development, 56,* 972-988. doi: 10.2307/1130109

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466,* 29. doi: 10.1038/466029a

Hughes, C. (2016). Theory of mind grows up: Reflections on new research on theory of mind in middle childhood and adolescence. *Journal of Experimental Child Psychology, 149,* 1-5. doi: 10.1016/j.jecp.2016/01.017

Kalish, C.W., & Lawson, C.A. (2008). Development of social category representations: Early appreciation of roles and deontic relations. *Child Development, 79,* 577-593. doi: 10.111/j1467-8624.2008.01144.x

Kramer, H.J. Goldfarb, D., Tashjian, S.M., & Lagattuta, K.H. (2017). "These pretzels are making me thirsty": Older children and adults struggle with induced-state episodic foresight. *Child Development, 88,* 1554-1562. doi: 10.1111/cdev.12700

Kramer, H.J., & Lagattuta, K.H. (in press). Developmental changes in emotion understanding during middle childhood. To appear in D.D. Dukes, E. Walle, & A.C. Samson, (Eds.), *Oxford handbook on emotional development.* Oxford University Press.

Kramer, H. J., Parra, L. A., Lara, K. H., Hastings, P. D., & Lagattuta, K. H. (2020). Consistency among social groups in judging emotions across time. *Emotion, 21,* 1-16. doi: 10.1037/emo0000836

Kushnir, T., Xu, F., & Wellman, H.M. (2010). Young children use statistical sampling to infer the preferences of others. *Psychological Science, 21*, 1134-1140. doi: 10.1177/0956797610376652

Lagattuta, K.H. (2005). When you shouldn't do what you want to do: Young children's understanding of desires, rules, and emotions. *Child Development, 76,* 713-733. doi: 10.1111/j.1467-8624.2005.00873.x

Lagattuta, K.H. (2008). Young children's knowledge about the influence of thoughts and emotions in rule situations. *Developmental Science, 11*, 809-818. doi: 10.1111/j.1467-7687.2008.00727.x

Lagattuta, K.H. (2014). Linking past, present, and future: Children's ability to connect mental states and emotions across time. *Child Development Perspectives, 8*, 90-95. doi: 10.1111/cdep.12065

Lagattuta, K. H. (2018). Where does it come from where does it go? The benefits of examining moral judgments across a wide age range. *Human Development, 60,* 350-356. doi: 10.1159/000485575

Lagattuta, K. H., Elrod, N. M., & Kramer, H. J. (2016). How do thoughts, emotions, and decisions align ? A new way to examine theory of mind in middle childhood and beyond. *Journal of Experimental Child Psychology, 149,* 116-133. doi: 10.1016/j.jecp.2016.01.013

Lagattuta, K.H., & Kramer, H.J. (2021). Advanced emotion understanding: Children's and adults' knowledge that minds generalize from prior emotion events. *Emotion, 21,* 1-16. doi: 10.1037/emo0000694

Lagattuta, K.H., & Kramer, H.J. (2021). Advanced theory of mind in middle childhood and adulthood. To appear in S. Lecce & R.T. Devine (Eds.), *Theory of mind in middle childhood and adolescence: integrating multiple perspectives*. Taylor and Francis.

Lagattuta, K.H., Kramer, H.J., Kennedy, K., Hjortsvang, K., Goldfarb, D., & Tashjian, S.M. (2015). Beyond Sally's missing marble: Further development in children's understanding of mind and emotion during middle childhood. *Advances in Child Development and Behavior, 48,* 185-217. doi: 10.1016/bs.acdb.2014.11.005

Lagattuta, K.H., & Sayfan, L. (2013). Not all past events are equal: Biased attention and emerging heuristics in children's past-to-future forecasting. *Child Development, 84,* 2094-2011. doi: 10.1111/cdev.12082

Lagattuta, K. H., & Thompson, R. A. (2007). The development of self-conscious emotions: Cognitive processes and social influences. In J. L. Tracy, R. W. Robins, & J. P. Tangney (Eds.), *The self-conscious emotions: Theory and research* (pp. 91-113). Guilford Press.

Lagattuta, K. H., Wellman, H. M., & Flavell, J. H. (1997). Preschoolers' understanding of the link between thinking and feeling: Cognitive cuing and emotional change. *Child Development, 68,* 1081-1104. doi: 10.2307/1132293

Lara, K.H., Lagattuta, K.H., & Kramer, H.J. (2019). Is there a downside to anticipating the upside? Children's and adults' reasoning about how prior expectations shape future emotions. *Child Development, 90,* 1170-1184. doi: 10.1111/cdev.12994

Lara, K. H., Kramer, H. J., & Lagattuta, K. H. (in press). This is not what I expected: The impact

of prior expectations on children's and adults' preferences and emotions. To appear in *Developmental Psychology.*

Ledgerwood, A., & Boydstun, A.E. (2014). Sticky prospects: Loss frames are cognitively stickier than gain frames. *Journal of Experimental Psychology: General, 143*, 376-385. doi: 10.1037/a0032310

Leppänen, U., Aunola K., Niemi, P., Nurmi, J. (2008). Letter knowledge predicts grade 4 reading fluency and reading comprehension. *Learning and Instruction, 18*, 548-564. doi: 10.1016/j.learninstruct.2007.11.004

Levin, I.P., & Gaeth, G.J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*, 374-378. doi: 10.1086/209174

Loewenstein, G., O'Donoghue, & Rabin, M. (2003). Projection bias in predicting future utility. *The Quarterly Journal of Economics, 118*, 1209-1248. doi: 10.1162/00335503322552784

Ma, L., & Xu, F. (2011). Young children' use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition, 120*, 403-411. doi: 10.1016/j.cognition.2011.02.003

Mahy, C.E.V. (2015). Young children have difficulty predicting future preferences in the presence of a conflicting physiological state. *Infant and Child Development, 25*, 325-338. Doi: 10.1002/icd.1930

Mahy, C.E.V., Grass, J., Wagner, S., Kliegel, M. (2014). These pretzels are going to make me thirsty tomorrow: Differential development of hot and cool episodic foresight in early childhood. *British Journal of Developmental Psychology, 32*, 65-77. doi: 10.1111/bjdp.12023

Mazachowsky, T.R., Koktavy, C., & Mahy, C.E.V (2019). The effect of psychological distance on young children's future predictions. *Infancy and Child Development, 28,* e2133. doi: 10.1002/icd.2133

McCormack, T., & Feeney, A. (2015). The development of the experience and anticipation of regret. *Cognition and Emotion, 29,* 266-280. doi: 10.1080/02699931.2014.912201

Milligan, K., Astington, J.W., & Dack, L.A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*, 622-646. doi: 10.1111/j.1467-8624.2007.01018.x

Nelson, N. L., & Russell, J. A. (2012). Children's understanding of nonverbal expressions of pride. *Journal of Experimental Child Psychology, 111*. 379-385. doi: 10.1016/j.jecp.2011.09.004

Nisbett, R.E., & Kanouse, D.E. (1969). Obesity, food deprivation, and supermarket shopping behavior. *Journal of Personality and Social Psychology, 12,* 289-294. doi: 10.1037/h0027799

Nunner-Winkler, G., & Sodian, B. (1988). Children's understanding of moral emotions. *Child Development, 59*, 1323-1338. doi: 10.2307/1130495

Pesowski, M.L., Denison, S., & Friedman, O. (2016). Young children infer preferences from a single action, but not if it is constrained. *Cognition, 155,* 168-175. doi: 10.1016/j.cognition.2016.07.004

Qualtrics (2021). Qualtrics Survey Software. Provo, Utah, USA. URL https://www.qualtrics.com

Quoidbach, J., Gilbert, D.T., & Wilson, T.D. (2013). The end of history illusion. *Science, 339,* 96-98. doi: 10.1126/science.1229294

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for

Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Read, D., van Leeuwen, B. (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational Behavior and Human Decision Processes, 76*, 189-205. doi: 10.1006/obhd.1998.2803

Roberts, S.O., & Gelman, S.A. (2016). Can White children grow up to be Black? Children's reasoning about the stability of emotion and race. *Developmental Psychology, 52*, 1830-1847. doi: 10.1037/dev0000132

Renoult, L., Kopp, L., Davidson, P.S.R., Taler, V., & Atance, C.M. (2016). You'll change more than I will: Adults' predictions about their own and others' future preferences. *Quarterly Journal of Experimental Psychology, 69*, 299-309. doi: 10.1080/17470218.2015.1046463

Repacholi, B.M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12-21. doi: 10.1037/0012-1649.33.1.12

Ruffman, T., Aitken, J., Wilson, A., Puri, A., Taumoepeau, M. (2017). A re-examination of the broccoli task: Implications for children's understanding of subjective desire. *Cognitive Development,* advanced online publication. doi: 10.1016/j.cogdev.2017.08.001

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science, 12*, 1123-1128. doi: 10.1177/1745691617708630

Thompson, R. A. (1987). Development of children's inferences of the emotions of others. *Developmental Psychology, 23*, 124-131. doi: 10.1037/0012-1649.23.1.124

Tillman, K.A., & Barner, D. (2015). Learning the language of time: Children's acquisition of duration words. *Cognitive Psychology, 78*, 57-77. doi: 10.1016/j.cogpsych.2015.03.001

Tracy, J. L., Robins, R. W., & Lagattuta, K. H. (2005). Can children recognize pride? *Emotion,*

5, 251-257. doi: 10.1037/1528-3542.5.3.251

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*, 371-384. doi: 10.1037/0033-295X.95.3.371

Wilson, T.D., & Gilbert, D.T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science, 14*, 131-134.

Weller, D., & Lagattuta, K.H. (2013). Helping the in-group feels better: Children's judgments and emotion attributions in response to prosocial dilemmas. *Child Development, 84*, 253-268. doi: 10.1111/j.1467-8624.2012.01837.x

Weller, D., & Lagattuta, K.H. (2014). Children's judgments about prosocial decisions and emotions: Gender of the helper and the recipient matters. *Child Development, 85*, 2011-2028. doi: 10.1111/cdev.12238

Wellman, H.M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Wellman, H.M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655-684. doi: 10.1111/1467-8624.00304

Wellman, H.M., Kushnir, T., Xu, F., & Brink, K.A. (2016). Infants use statistical sampling to understand the psychological world. *Infancy, 21*, 668-676. doi: 10.1111/infa.12131

Chapter 2

Consistency Among Social Groups in Judging Emotions Across Time

Abstract

We measured judgments about emotions across time. In Study 1 (N = 254) and Study 2 (N = 162), LGBTQ-Latinx, Straight-Latinx, LGBTQ-White, and Straight-White emerging adults rated how they would feel if a perpetrator acted positively (P) or negatively (N) towards them in single, isolated events. In Study 2, participants also responded to a new *Emotions Across Time Task* where they judged how they would feel interacting with a hypothetical perpetrator across three timepoints: (1) an initial past event, (2) a recent past event, and (3) an uncertain future-oriented event (seeing the perpetrator again). Participants further predicted their thoughts and decisions in the uncertain future-oriented event. The past emotional events appeared in various sequences (PP, NN, NP, PN). Results indicated that participants judged events as unambiguous when occurring first in a sequence or in isolation (positive events feel better than negative events). In contrast, initial events shaped emotional reactions to subsequent events: Participants responded more intensely to episodes that were preceded by events of the same valence. In addition to this augmenting effect, initial negative events were especially sticky: Participants rated a positive event following a negative event as feeling less good than when a positive event appeared first or in isolation, but they judged negative events to feel equivalently bad regardless of order. When evaluating future-oriented affective states, participants drew from the prior experiences and prioritized the recent past (more positive emotions, thoughts, and decisions for PP > NP > PN > NN). Effects replicated across all social groups.

Key words: Future thinking, emotion understanding, social cognition, marginalized social groups, LGBTQ, Latinx

Consistency Among Social Groups in Judging Emotions Across Time

Reactions to emotional events often seem unambiguous. For example, most people would judge that receiving a desirable gift feels good. Although this general, script-based knowledge can get emotion-perceivers and emotion-experiencers quite far in predicting, explaining, and understanding emotions, it is important to also recognize that emotional episodes rarely occur in isolation. Rather, a variety of outside forces shape affective reactions to events. Returning to the prior example, imagine that the gift-giver had previously harmed you—does receiving the gift still feel as rewarding? In contrast, what if the gift-giver had previously done another generous act towards you, does receiving the current gift now feel more intensely positive? In the current research, we tested lay intuitions about the shaping power of the past on subsequent emotional responding. We created an *Emotions Across Time Task* (*EATT*) to assess whether adults expect prior life experiences to influence later affective responses to positive and negative events as well as to uncertain future-oriented situations. While our focus was on how adults (as a general group) think about emotions across time, we considered whether belonging to a marginalized social group is related to judgments about how past events influence present reactions to current and future-orientated situations, particularly in ambiguous social contexts (Inzlicht et al., 2009). Thus, we included groups who have historically been marginalized because of their sexual (i.e., people who self-identify as lesbian, gay, bisexual, transgender, and queer; LGBTQ[9]; Meyer, 1995, 2003) and/or ethnic identities (i.e., Latinx[10] people; Paradies, 2015).[11]

[9]Although sexual orientation and gender are considered distinct social constructs, these social group memberships are often grouped together under the coalition acronym of LGBTQ (lesbian, gay, bisexual, transgender or queer; Institute of Medicine, 2011). In this study, we use the term LGBTQ to be inclusive, while specifically focusing on sexual orientation and acknowledging distinct differences between these groups under the LGBTQ coalition term.
[10]The term Latinx transcends the male versus female, sex-gender binary that is inherent in the Spanish language to be inclusive of all sexually and gender diverse people of Latin American descent (Scharrón-del Río & Aja, 2015).
[11]This paper is a published manuscript. Citation: Kramer, H. J., Parra, L. A., Lara, K. H., Hastings, P. D. (2020). Consistency among social groups in judging emotions across time. *Emotion, advanced online*. doi: 10.1037/emo000836

When reasoning about emotions, adults exhibit awareness that emotions do not simply arise from features of a current situation. For example, adults believe that holding previously low expectations (versus high expectations) leads to more positive emotions after an outcome is known (Lara et al., 2019; Shepperd & McNulty, 2002). They also appreciate that thinking about how things could have been better leads to more negative feelings than if the alternative had not been considered; likewise, imagining how an outcome could have been worse improves emotional wellbeing (Atkinson et al., 2009; Payir & Guttentag, 2016; Roese, 1997). More generally, adults understand that people's thoughts and interpretations of a situation can bias their affective responses (Kramer & Lagattuta, in press; Lagattuta, 2014; Lagattuta et al., 2015). Here, we investigated beliefs about another potential emotion elicitor outside of the immediate situation. We assessed whether adults consider prior life experiences—the events that preceded the current situation—as viable influencers of how individuals will respond to present negative and positive events. Episodes following events of the opposite valence (i.e., a negative event occurring after a positive event; a positive event happening after a negative event) may feel less intense because they are colored by the initial experience. In contrast to this emotion dampening, emotional events of the same valence occurring in sequence could be augmented. That is, potentially, a second negative episode feels even worse than an initial negative event, whereas a subsequent positive event following an initial positive event could feel even more exhilarating.

When considering how the past influences current emotional reactions, negative events may be stickier than positive events. Previous research indicates that people have a negativity bias in several domains (Baumeister et al., 2001; Vaish et al., 2008). For example, children and adults show a natural tendency to attend to negative information, even when directed to look only at positive stimuli (Lagattuta & Kramer, 2017; LoBue & DeLoache, 2008; Öhman et al.,

2001). Children also exhibit more sophisticated reasoning about negative compared with positive emotions (Bamford & Lagattuta, 2012; Lagattuta & Wellman, 2001; Lagattuta & Wellman, 2002; Lagattuta et al., 1997; Lara et al., 2019). As well, when information is initially framed negatively (e.g., a new jobs program will lose 40% of jobs), adults' attitudes shift less once that information is subsequently reframed positively (e.g., that means that the new jobs program will save 60% of jobs) compared with when the order of frames is reversed (Boydstun et al., 2018; Ledgerwood & Boydstun, 2014; Sparks & Ledgerwood, 2018). Relatedly, adults require more evidence to judge that someone with negative traits (e.g., selfish, unfriendly) has developed positive traits (e.g., selfless, friendly) than it takes for them to believe the inverse progression (Klein & O'Brien, 2016; O'Brien & Klein, 2017). In the current study, we tested whether adults judge that negative past events bias reactions to subsequent positive events more strongly than do past positive events shape reactions to later negative episodes.

Although no prior work has assessed lay theories about the potential dampening or augmenting effects of prior events on a person's emotional response to a current outcome, researchers have tested children's and adults' beliefs about how prior life experiences bias affective responses to future-oriented situations. Lagattuta and Sayfan (2013) showed 4- to 10-year-olds and adults scenarios in which a perpetrator caused a focal child to feel negative or positive emotions on two occasions in varying orders: negative followed by negative (NN), positive followed by positive (PP), negative followed by positive (NP), and positive followed by negative (PN). In a final scene, the focal character sees the perpetrator again and participants judged the focal characters' thoughts (whether the character thought something good or bad would happen next), emotions (whether the character felt happy or worried), and decisions (whether the character would approach or avoid the perpetrator). Children and adults provided

more positive emotion ratings, expected a higher likelihood of a positive future, and provided closer approach decisions for PP > NP > PN > NN pasts. The reliance on past event information, however, increased within childhood as well as between childhood and adulthood (see also Lagattuta, 2007; Lagattuta & Kramer, 2019; Lagattuta et al., 2018).

These distinctions among past types when predicting future-oriented mental states (i.e., more positive attributions for PP > NP > PN > NN) reveal that participants across a wide age range believe that the past matters. What is less clear is how individuals incorporate and weight each prior episode. Lagattuta and Sayfan (2013) argued for a recency bias: Children and adults attributed more positive reactions following NP versus PN pasts, and they also visually attended to pictorial stimuli depicting the recent past more than the initial past (especially when the recent past was negative; assessed via eye tracking). In contrast, other work has shown that adults rely on initial information when forming impressions (Asch, 1946; Gawronski & Bodenhausen, 2006; Uleman & Kressel, 2013), suggesting that future reactions should be anchored to the first event. These perspectives may not be at odds. That is, perhaps a primacy and recency bias jointly operate. Lagattuta and Sayfan (2013) told and showed participants how characters felt after each past event and had pre-matched the intensity of negative and positive events in NP and PN trials based on pilot participants' ratings for each event in isolation. It is possible that had participants actually evaluated characters' emotions after each event in sequence during the paradigm, they may have reasoned differently about how characters felt in the recent past and about the future. That is, as we conjectured above, participants may have expected characters' reactions to the second past episode to be biased by what had happened first (e.g., a positive event following a negative event is not as positive as that same positive event in isolation). Thus, we examined the extent to which adults incorporate both primacy and recency biases in their affective judgments

about current events (a positive or negative event presently occurring) as well as when reasoning about future-oriented events (anticipating what will happen next).

Present Research

We conducted two studies to assess emotional reactions to events across time. In Studies 1 and 2, adults rated how they would feel experiencing negative and positive events in isolation. In Study 2, we modified the past-to-future measure from Lagattuta and Sayfan (2013) to create an Emotions Across Time Task (EATT). Participants provided emotion ratings for each of the past events in the sequence prior to judging future-oriented reactions. This enabled us to test beliefs about (1) whether and how an *initial event* influences emotional reactions to a *recent (subsequent) event* (e.g., by comparing ratings to the same positive event when it appeared first versus second in a sequence; and, when second if preceded by a positive or negative initial event), as well as (2) whether and how two emotional past events bias emotions, thoughts, and decisions in an *uncertain future-oriented event* (i.e., seeing that same past perpetrator at a later time point). By having emotion ratings at all three time points (initial event, recent event, future-oriented event), we could further test which past event(s) adults relied on most (initial, recent, average emotion rating across the past event sequence) when reasoning about the future.

In the second experiment, we also address some additional critical questions that remain unanswered from Lagattuta and Sayfan (2013), especially with regard to the adult response patterns. In particular, because their task was primarily designed to address developmental changes in past-to-future reasoning, Lagattuta and Sayfan (2013) created a highly structured paradigm involving illustrations and narrations to aid in comprehension. Furthermore, all participants reasoned about child protagonists. Thus, adult reasoning could have been driven by them responding to how they believe children would think, feel, and make decisions in these

situations, not how they think the past influences future-oriented responding more generally. Moreover, adults may also think differently if they are asked to consider their own, first-person reactions rather than reason about how other people will feel. Thus, in the current study we further tested whether adults' beliefs about how past experiences influence future-oriented affective reactions documented by Lagattuta and Sayfan (2013) would replicate when the task is stripped down to non-pictorial, first-person, adult-relevant incidents (i.e., PP > NP > PN > NN).

In both studies, we included representation of people from marginalized groups. We intentionally recruited individuals belonging to one or more marginalized social groups to examine how this status influences judgments about emotional reactions to events in isolation, events in sequence, and how past events influenced responses to ambiguous future-oriented events. Theoretical frameworks suggest that marginalized people have a more pronounced negativity bias when navigating ambiguous or threatening social contexts (Crocker et al., 1998; Goffman, 1963; Jones et al., 1984). This bias may aid in the detection of negative social events or threats that could be perceived as discrimination (Inzlicht et al., 2009). Studies show that LGBTQ and Latinx people anticipate future discriminatory experiences based on the intersection of their marginalized sexual and ethnic group memberships (Scheim & Bauer, 2019). Yet, little is known about the on-line emotion cognition that may be associated with how LGBTQ, Latinx, and LGBTQ-Latinx people anticipate future events based on past experiences.

Because we recruited individuals with multiple marginalized social group memberships (e.g., LGBTQ-Latinx), we used an intercategorical quantitative application (McCall, 2005) of the intersectionality framework (Collins, 1991; Crenshaw, 1989) to inform the composition of our groups for analyses. For this reason, for analyses we grouped LGBTQ-Latinx, LGBTQ-White, Straight-Latinx, and Straight-White people into their own intersectional social groups (similar to

Bauer & Scheim, 2019; Scheim & Bauer, 2019) to acknowledge: (1) that sexual orientation and ethnicity are tightly interwoven social constructs (Garnets, 2002); (2) that LGBTQ people of color's ethnic or racial backgrounds influence the meanings they ascribe to their sexual identities (DeBlaere et al., 2010); and (3) that people vary in their experiences and perceptions of negative events at various sexual and ethnic social group intersections (Bauer & Scheim, 2019; Scheim & Bauer, 2019). Thus, we made this analytical decision instead of adhering to a more traditional between-groups (2 X 2; sexual orientation x ethnicity) study design because using the standard approach is an oversimplified proxy for how the identities of people who belong to multiple marginalized social groups intersect (Cole, 2009; Crenshaw, 1989; Parent et al., 2013).

Hypotheses

In Studies 1 and 2, we hypothesized that when in isolation, events we *a priori* determined to be positive would be rated more positively than those that we *a priori* determined to be negative. We expected that negative events would be rated as more intensely negative by marginalized social groups than by non-marginalized social groups. In Study 2, we predicted that adults would expect initial past negative and positive events to influence emotional reactions to subsequent negative and positive events, with initial negative episodes causing a stronger bias than initial positive episodes, particularly for marginalized social groups. We also anticipated that adults would provide more intensely positive emotions, thoughts, and decisions upon seeing agents of PP > NP > PN > NN pasts (conceptually replicating Lagattuta & Sayfan, 2013), and we explored whether these effects would be moderated by social group membership. To assess the assumption that adults rely most on the recent past when thinking about the future, we compared participants' future-oriented emotion ratings to the emotion ratings they provided for the initial past event, the recent past event, and their average emotion rating across the past event sequence.

Study 1

Method

Participants included two hundred and fifty-four emerging adults between the ages of 18 and 29 years (M = 22.71 years; SD = 3.26) in four self-identified groups: LGBTQ-Latinx (n = 63), Straight-Latinx (n = 66), LGBTQ-White (n = 58), and Straight-White (n = 67).[12] Our sample size was based on prior work on emotion cognition (Lagattuta & Sayfan, 2013). Participants were eligible if they were not incarcerated; identified as Latinx or White; identified as female, male, transgender, genderqueer, or gender non-conforming; could speak and comprehend English fluently; and if they were between 18 and 29 years of age. Participants were recruited through a university subject pool, social media, and through emails to listservs for LGBTQ and Latinx community and student groups (*see Table B.1* for additional demographics). All participants were entered in a raffle to win one of eight $50 gift cards. Seventeen participants did not answer enough of the items to calculate the DVs, and were excluded from analyses (final N = 237; 58 LBGTQ-Latinx, 65 Straight-Latinx, 52 LGBTQ-White, 62 Straight-White). This study was approved by the Institutional Review Board at University of California, Davis, #1122593-2.

Events in Isolation

Participants read 36 events in which a "perpetrator" hypothetically acted positively (n = 17 events; e.g., "Someone praised you") or negatively (e.g., "Someone rejected you"; n = 19 events) towards the participant (*Table B.2*). For each event, participants reported the valence of the event (7-point scale from very negative to very positive). Participants responded to the events in random order. These 36 events were informed by questionnaire items from scales that assess

[12]A greater number of people (N = 270) consented to being part of the study of which (n = 7) did not complete any questionnaire data, (n = 4) did not provide information about their sexual orientation, and (n = 5) did not identify as Latinx or White. The cases pertaining to these respondents were not included in the analyses.

general daily hassles (Brantley et al., 2007), ethnic discrimination (Brondolo et al., 2005), and sexual discrimination (Rosario et al., 2002). To ensure that the final 36 events would be appropriate to use with all social groups (e.g., Scheim & Bauer, 2019), none of the episodes were specific to experiences related to sexual orientation or ethnicity, and none explicitly described the perpetrator's actions as influenced by sexual orientation or ethnic social group membership.

We coded participants' valence ratings on a 7-point scale: -3 = Very Negative; -2 = Medium Negative; -1 = A Little Negative; 0 = Neutral (not negative or positive); 1 = A Little Positive; 2 = Medium Positive; 3 = Very Positive. We averaged across event type for primary analyses to calculate an average emotion rating for positive events and an average emotion rating for negative events (see *Table B.2* for means and standard deviations by each individual event). During averaging, if participants were missing one or more items, we calculated their average out of the total number of events to which they did respond (e.g., if a participant only responded to 16 of the 19 negative events, then her average was calculated out of 16 rather than 19 events).

General Procedure

Participants first answered questions regarding their eligibility for the study. Next, eligible participants provided informed consent. We collected all data (including eligibility and consent) via Qualtrics Survey Software (Qualtrics, 2019). Prior to responding to the valence rating task, participants provided demographic information (including sexual orientation and ethnic group membership). Next, participants responded to the emotion valence rating measure. Within this measure, participants also reported how frequently each event had happened to them in the past, how often they expected each event to happen to them in the future, and how common they thought these events were in other people's lives. As well, participants reported the specific discrete emotion or emotions they expected to accompany each event (sad, mad,

worried/scared, ashamed, disappointed, OK/neutral, happy, comforted, excited, proud, relieved). After completing the event ratings, participants reported on their general beliefs and emotional experiences. We describe these additional measures to be transparent in our reporting. These surveys and tasks, however, will be analyzed in separate manuscripts. At the end of the survey, we debriefed participants and they were invited to enter the raffle.

Results and Discussion

Analyses were conducted in RStudio (RStudio Team, 2016). We set alpha = .05.

Events in Isolation Task (Table 5, Figure 5)

We conducted a 4 (social group membership: LGBTQ-Latinx, Straight-Latinx, LGBTQ-White, Straight-White) x 2 (valence: negative, positive) repeated measures analysis of variance (ANOVA) on valence ratings.[13] This analysis resulted in a main effect for valence, $F(1, 233) = 3829.97$, $p < .001$, $\eta_p^2 = .94$, but no significant effects for social group membership, Fs < 1.85, ps > .139, η_p^2s < .02. As predicted, participants provided more intensely positive valence ratings following positive events than after negative events. Put more simply, participants rated positive events as about "Medium Positive" and negative events as approximately "Medium Negative."

[13]When we conducted analyses as a 2 (sexual orientation) x 2 (ethnicity) design, the same patterns emerged.

Table 5. Chapter 2: *Study 1 and Study 2 Means and Standard Deviations for Events in Isolation (Emotion Ratings), Events in Sequence (Emotion Ratings), and Past-to-Future by Social Group Membership.*

	LGBTQ-Latinx	Straight-Latinx	LGBTQ-White	Straight-White	All Participants
			Study 1		
			Events in Isolation		
Positive (P)	2.17 (0.58)	2.18 (0.52)	1.99 (0.47)	1.99 (0.82)	2.09 (0.62)
Negative (N)	-1.88 (0.52)	-1.80 (0.58)	-1.86 (0.46)	-1.77 (0.52)	-1.83 (0.53)
			Study 2		
			Events in Isolation		
Positive (P)	2.16 (0.69)	2.14 (0.52)	1.95 (0.47)	2.08 (0.52)	2.09 (0.56)
Negative (N)	-1.94 (0.57)	-1.96 (0.51)	-2.12 (0.42)	-2.04 (0.47)	-2.01 (0.50)
			Events in Sequence		
PP Initial (P)	2.03 (0.71)	2.01 (0.72)	1.79 (0.55)	2.09 (0.59)	1.99 (0.65)
PP Recent (P)	2.27 (0.80)	2.19 (0.68)	2.31 (0.45)	2.28 (0.52)	2.26 (0.63)
PN Initial (P)	2.02 (0.81)	2.06 (0.80)	1.92 (0.71)	2.00 (0.77)	2.00 (0.77)
NP Recent (P)	0.44 (1.30)	0.57 (1.02)	0.50 (1.13)	0.90 (0.96)	0.61 (1.12)
NN Initial (N)	-1.99 (0.85)	-1.93 (0.71)	-1.97 (0.69)	-1.99 (0.64)	-1.97 (0.72)
NN Recent (N)	-2.40 (0.63)	-2.39 (0.62)	-2.44 (0.61)	-2.60 (0.49)	-2.46 (0.59)
NP Initial (N)	-1.94 (0.71)	-1.83 (0.66)	-2.00 (0.64)	-1.96 (0.63)	-1.94 (0.66)
PN Recent (N)	-1.91 (0.76)	-1.94 (0.78)	-2.03 (0.63)	-2.09 (0.65)	-1.99 (0.71)
			Past-to-Future		
PP Emotion	2.33 (0.62)	2.23 (0.79)	2.40 (0.56)	2.35 (0.54)	2.33 (0.63)
NP Emotion	-0.15 (1.19)	-0.41 (1.05)	-0.56 (1.29)	-0.05 (1.00)	-0.28 (1.14)
PN Emotion	-1.32 (1.20)	-1.59 (0.88)	-1.60 (1.10)	-1.54 (0.80)	-1.50 (1.01)
NN Emotion	-2.23 (0.81)	-2.39 (0.72)	-2.44 (0.54)	-2.46 (0.57)	-2.38 (0.68)
PP Thought	2.02 (0.54)	1.99 (0.66)	2.03 (0.53)	2.13 (0.62)	2.04 (0.59)
NP Thought	-0.28 (1.26)	-0.10 (1.12)	-0.15 (1.32)	0.10 (0.94)	-0.11 (1.16)
PN Thought	-0.82 (1.01)	-0.93 (0.67)	-0.93 (0.96)	-0.65 (0.93)	-0.82 (0.94)
NN Thought	-2.18 (0.76)	-2.20 (0.70)	-2.24 (0.72)	-2.17 (0.53)	-2.20 (0.68)
PP Decision	2.33 (0.59)	2.29 (0.81)	2.38 (0.57)	2.61 (0.48)	2.40 (0.62)
NP Decision	-0.16 (1.52)	-0.36 (1.35)	-0.26 (1.51)	-0.10 (1.30)	-0.21 (1.41)
PN Decision	-0.75 (1.28)	-1.00 (1.33)	-0.85 (1.49)	-0.87 (1.16)	-0.86 (1.30)
NN Decision	-2.45 (0.96)	-2.40 (1.10)	-2.51 (0.90)	-2.54 (0.76)	-2.48 (0.93)

Note. Valence Ratings for Study 1: -3 = Very Negative; -2 = Medium Negative; -1 = A Little Negative; 0 = Neutral (Not Negative or Positive); 1 = A Little Positive; 2 = Medium Positive; 3 = Very Positive; Valence Ratings for Study 2: -3 = Very Bad; -2 = Medium Bad; -1 = A Little Bad; 1 = A Little Good; 2 = Medium Good; 3 = Very Good. "Initial" = emotion rating for first event in the sequence; "Recent" = emotion rating for second event in the sequence.

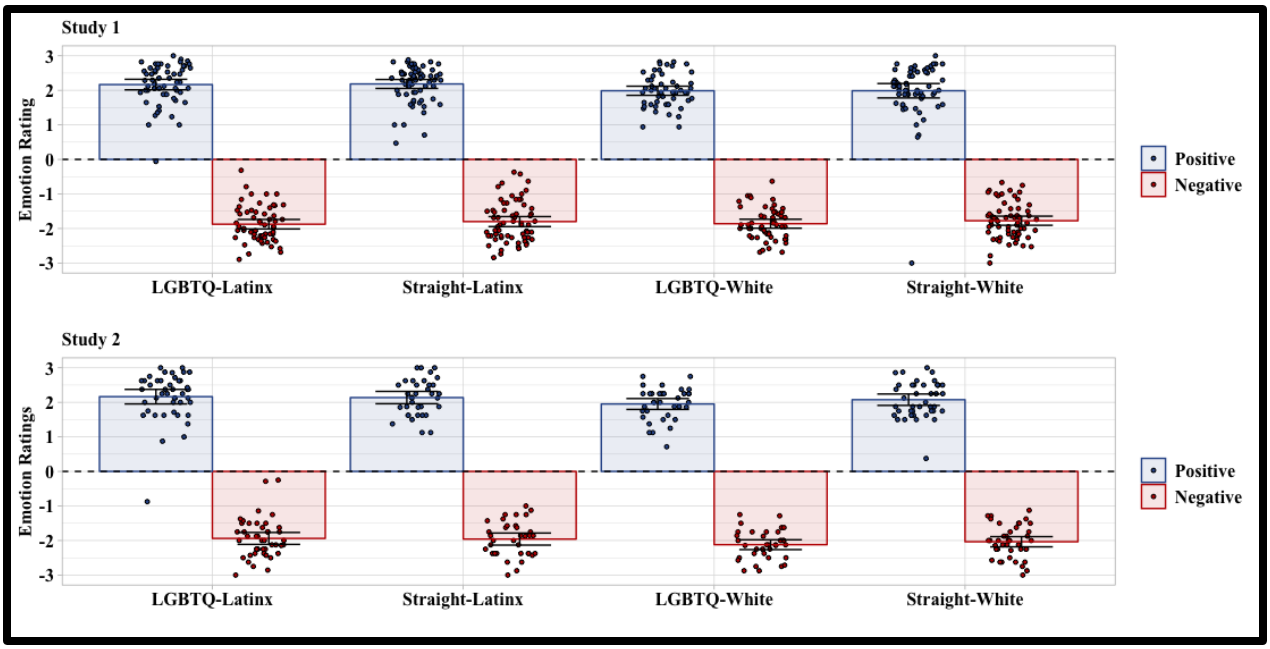

Figure 5. Chapter 2: Events in isolation ratings by study, group, and valence. Bar = Mean; Error bar = 95% CI; small circle = jittered individual data.

These data reveal that adults share consistent beliefs about emotional reactions to negative and positive events in isolation. Events that we assumed would be rated positively were rated more positively than events that we predicted would be rated negatively. A secondary goal of Study 1 was to assess whether there were any social group differences in how people reason about the impact of negative and positive events on their emotions. We found no significant group differences: Participants rated negative events as negatively and positive events as positively regardless of their social group membership.

Study 2

The aim of Study 2 was to measure beliefs about how events from the past influence emotional reactions to current events (i.e., the participant imagines someone doing something bad or good to them after this person has already done something bad or good to them) and uncertain future-oriented events (i.e., the participant imagines seeing someone who has previously done good, bad, or both good and bad actions to them in the past). For example, do

adults anticipate feeling more positively receiving a desirable gift from someone if that person had previously praised them versus called them a derogatory word? Furthermore, does seeing someone from the past feel better if that person previously rejected you and then later celebrated your accomplishments compared with if they first celebrated your accomplishments and then later rejected you? We again examined potential differences by social group membership. Although we found no evidence for group differences in Study 1, it is possible that when considering emotional events in sequence that group differences emerge. For example, perhaps initial negative events are stickier for marginalized groups (e.g., if someone has done something bad in the past, nothing can make up for such an event).

Method

Participants

Participants included 162 adults: 45 LGBTQ-Latinx, 38 Straight-Latinx, 36 LGBTQ-White, and 43 Straight-White emerging adults between the ages of 20 and 31 years (M = 25.15 years; SD = 2.60). These data were collected at a third time point of a longitudinal study assessing the impact of discrimination on mental and physical health in sexually and ethnically diverse people over the course of two years.[14] Participants were recruited through social media, LGBTQ and Latinx community and student groups, as well as at Pride month events in Davis and Sacramento, California (see *Table B.3* for additional demographics). Six participants did not answer enough of the items to calculate the DVs, and were excluded from analyses (final N = 156; 44 LGBTQ-Latinx, 35 Straight-Latinx, 36 LGBTQ-White, 41 Straight-White). Because Study 2 was part of longitudinal study, we excluded people from Study 1 who reported being

[14]In the parent longitudinal study, the sample sizes at first and second waves were (N = 202; 51 LGBTQ-Latinx, 49 Straight-Latinx, 51 LGBTQ-White, and 51 Straight-White; M = 23.13 years; SD = 2.59) and (N = 171; 45 LGBTQ-Latinx, 40 Straight-Latinx, 42 LGBTQ-White, and 44 Straight-White; M = 23.99 years; SD = 2.57), respectively.

part of this particular longitudinal study. This study was approved by the Institutional Review Board at University of California, Davis, #832712-14.

Event Selection

We used participants' ratings from Study 1 to select 8 negative and 8 positive items for use in Study 2 (see *Table B.4*). All selected negative events had valence ratings between approximately "Medium Negative" and "Very Negative" (-2.43 < *M*s < -1.82; 0.80 < *SD*s < 1.14). All positive events had valence ratings between about "Medium Positive" and "Very Positive" (1.98 < *M*s < 2.44; .74 < *SD*s < 1.31). On average, negative events were within 0.28 scale points of each other, positive events were within 0.22 scale points of each other, and positive events were within 0.24 scale points from negative events (in terms of intensity).

Events in Isolation Task

Participants reported how they would feel after 8 negative events (e.g., "One day, a person you have never met before *damaged your property*.") and 8 positive events (e.g., One day, a person you have never met before *praised you*.") on a 6-point scale from Very Bad to Very Good. On each trial, participants were instructed to imagine someone new who they had never met before. The order of events was randomized. We coded participants' emotion ratings in isolation on a 6-point scale: -3 = Very Bad; -2 = Medium Bad; -1 = A Little Bad; 1 = A Little Good; 2 = Medium Good; 3 = Very Good. For analyses, we averaged across the eight negative events to create one negative rating and the eight positive events to create one positive rating. We handled missing data in the same way as Study 1.

Emotions Across Time Task (EATT)

Participants then responded to a series of two-event sequences. During the first event, the perpetrator did something negative (e.g., "One day, a person you have never met before *ignored*

your pleads for help.") or positive (e.g., "One day, a person you have never met before *tried to cheer you up.*"). The participant reported how they would feel using the same emotion scale described in the *events in isolation* task. Next, participants imagined this same person doing something negative (e.g., "A few days later, this same person *rejects you.*") or positive (e.g., "A few days later, this same person *praises you.*"), and they reported how they would feel.

Participants then imagined seeing the same person again (e.g., "Remember, these two things happened to you. First, this person *rejected you*. A few days later, this person *praised you.* Many days later, you see this same person again."). Participants reported how they would feel (using the same emotion scale), what they thought the person would do next (6-point scale: from "Definitely will do something bad" to "Definitely will do something good"), and what the participant would do next (6-point scale: "Really sure I would stay away from this person" to "Really sure I would go near this person"). Participants were also asked to explain why they would make the decision that they did. Consistent with previous work (Lagattuta & Kramer, 2019; Lagattuta & Sayfan, 2013), we reminded participants of both past events (initial and recent) immediately prior to asking them to report their emotions, thoughts, and decisions. This recap removed memory constraints and ensured that both episodes were made equally salient.

Participants responded to eight of these 2-event sequences in varying order: negative then negative (NN; two trials), positive then positive (PP; two trials), negative then positive (NP; two trials), and positive then negative (PN; two trials). For each sequence, participants were instructed to imagine someone new whom they had never met before. All of the events used in the *events in isolation task*, were also used in the events in sequence and past-to-future reasoning tasks. The task was programmed such that each of the events could be slotted in to any sequence (with the constraint that participants only saw each event once during the events in sequence and

past-to-future reasoning task). Specific sequences and order of sequences were randomized.

We coded emotion ratings to each of the two events in sequence as well as to seeing the perpetrator again on a 6-point scale: -3 = Very bad; -2 = Medium bad; -1 = A little bad; 1 = A little good; 2 = Medium good; 3 = Very good. We coded thought ratings on a 6-point scale: -3 = Definitely will do something bad, -2 = Probably will do something bad, -1 = Might do something bad, 1 = Might do something good, 2 = Probably do something good, 3 = Definitely will do something good. We coded decision ratings on a 6-point scale: -3 = Really sure I would stay away from this person, -2 = Kind of sure of would stay away from this person, -1 = Not so sure I would stay away from this person, 1 = Not so sure I would go near this person, 2 = Kind of sure I would go near this person, 3 = Really sure I would go near this person. We averaged across the two trials from each past type. Missing data were dealt with the same way as Study 1.

General Procedure

We collected all data (including consent) via Qualtrics Survey Software (Qualtrics, 2019). Because this was a longitudinal study, eligibility was assessed during the first wave of the study, and was not reassessed here. After providing consent, participants completed demographic information (including sexual orientation and ethnic group membership). Next, participants completed the events in isolation task followed by the EATT. After both tasks, participants were shown the emotional events one more time and they reported how frequently the events had happened to them in the past, how often they expected them to happen to them in the future, and how common these events were in other people's lives. Participants then completed mental health measures. These individual difference measures will be analyzed in separate manuscripts. Participants were debriefed and compensated with a $25 gift-card for their time and efforts.

Results and Discussion

Results are presented in four sections. We first analyze beliefs about emotional events in isolation. Next, we analyze responses to the EATT, separating judgments for emotional reactions to positive and negative events occurring in varying sequences (PP, NP, PN, NN) and reasoning about the influence of past event sequences on future-oriented affective responses. Finally, we analyze how participants differentially weight prior events when thinking about the future. Analyses were conducted in RStudio (Rstudio, 2016). For all analyses, we set alpha = .05, and we used Tukey's HSD to correct for multiple comparisons.

Events in Isolation (Table 5, Figure 5)

We conducted a 4 (social group membership) x 2 (valence) repeated measures ANOVA on average valence ratings. This analysis resulted in a main effect for valence, $F(1, 152) = 4110.46$, $p < .001$, $\eta_p^2 = .96$, and no effects for social group membership, Fs < 2.37, ps > .073, η_p^2s < .04. As expected, and replicating Study 1, participants provided more intensely positive emotion ratings following positive events than after negative events. Indeed, similar to Study 1, participants reported that the positive events felt approximately "Medium Good" and negative events felt about "Medium Bad."

EATT: Emotions in Sequence (Table 5, Figure 6)

We conducted a 4 (social group membership) x 4 (past: PP, NP, PN, NN) x 2 (event: initial, recent) repeated measures ANOVA on emotion ratings. This analysis yielded a main effect for past, $F(3, 456) = 1611.68$, $p < .001$, $\eta_p^2 = .91$, and event, $F(1, 152) = 134.18$, $p < .001$, $\eta_p^2 = .47$, qualified by a Past x Event interaction, $F(9, 456) = 1153.66$, $p < .001$, $\eta_p^2 = .88$. There were no effects for social group membership, Fs < 1.07, ps > .387, η_p^2s < .02. As would be

expected, initial events were treated equivalently to events in isolation.[15] That is, when evaluating how they would feel after initial events, participants rated the two negative events equivalently (the initial N in *N*N and *N*P; $p > .999$) and the two positive events equivalently (the initial P in *P*P and *P*N; $p > .999$). Furthermore, they reported that the two initial negative events would feel worse than the two initial positive events ($ps < .001$).

As predicted, emotions in response to the second event were biased by the initial event. A positive event preceded by a positive event (P*P*) was rated more positively than a positive event that occurred first in a sequence (*P*P, *P*N, $ps < .035$). Moreover, a negative event preceded by a negative event (N*N*) was rated more negatively than when a negative event occurred first in a sequence (*N*N, *N*P, $ps < .001$). Consistent with the heightened stickiness of negative information over positive information, a positive event that came after a negative event (N*P*) was rated less positively than when it appeared first (*P*P, *P*N, $ps < .001$), but a negative event that followed a positive event (P*N*), was rated as negatively as an initial negative event (*N*N, *N*P, $ps > .996$).

[15]When comparing the two initial negative events (***N***N, ***N***P) to the negative event in isolation there were no effects for event, $Fs < 1.26$, $ps > .286$, $\eta_p^2 s < .01$. Similarly, when comparing the two initial positive events (***P***P, ***P***N) to the positive event in isolation, there were no effects for event, $Fs < 2.60$, $ps > .076$, $\eta_p^2 s < .02$.

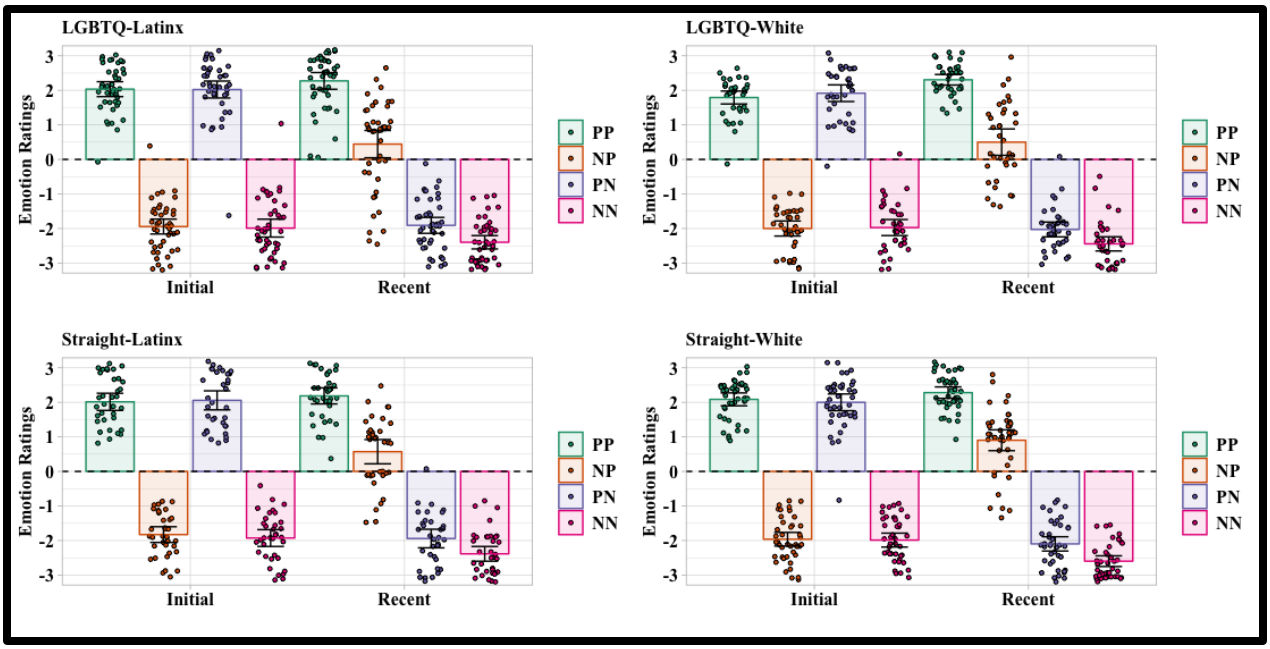

Figure 6. Chapter 2: Events in sequence ratings by group, past, and event. PP = Positive then Positive; NP = Negative then Positive; PN = Positive then Negative; NN = Negative then Negative. Bar = Mean; Error bar = 95% CI; small circle = jittered individual data. "Initial" = emotion rating for first event in the sequence; "Recent" = emotion rating for second event in the sequence.

EATT: Past-to-Future Reasoning (Table 5, Figure 7).

Three separate 4 (social group membership) x 4 (past) repeated measures ANOVA on (a) emotion intensity ratings, (b) thought likelihood ratings, and (c) decision certainty ratings all resulted in a main effect for past (emotion: $F[3, 456] = 899.83$, $p < .001$, $\eta_p^2 = .86$; thought: $F[3, 456] = 713.34$, $p < .001$, $\eta_p^2 = .82$; decision: $F[3, 456] = 619.12$, $p < .001$, $\eta_p^2 = .80$), and no effects for social group membership (emotion: Fs < 1.56, ps > .201, η_p^2s < .03; thought: Fs < 1.13, ps > .339, η_p^2s < .02; decision: Fs < 0.41, ps > .812, η_p^2s < .01). Conceptually replicating and extending Lagattuta and Sayfan (2013), participants anticipated that a positive future felt better, was more likely, and made more confident approach decisions upon seeing a perpetrator of PP > NP > PN > NN pasts (ps < .001).

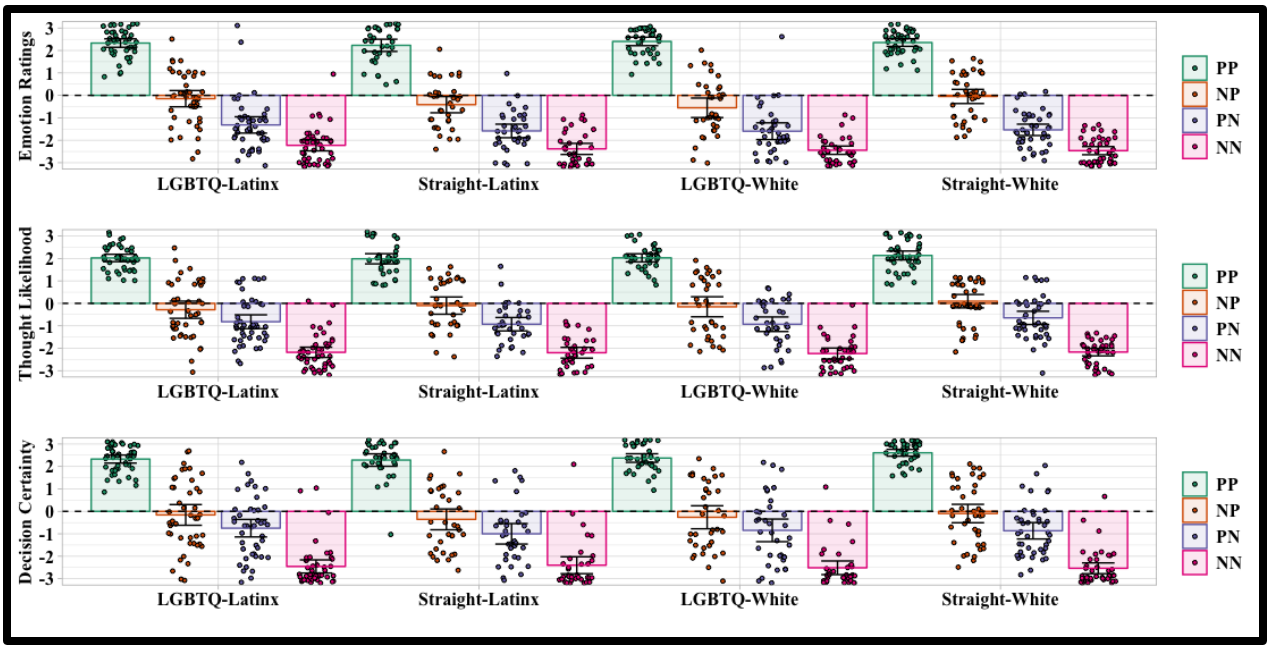

Figure 7. Chapter 2: Past-to-future ratings by judgment, group, and past. PP = Positive then Positive; NP = Negative then Positive; PN = Positive then Negative; NN = Negative then Negative. Bar = Mean; Error bar = 95% CI; small circle = jittered individual data.

Comparing Future-Oriented Emotions to Past Emotional Reactions (Table 6).

We created difference scores between the emotion rating for the future-oriented event and the (1) initial past event (Event 1), (2) recent past event (Event 2), and (3) past average emotion rating (see *Table 6* for means and standard deviations, including information about the direction of the difference). Using these scores, we compared participants' future-oriented reactions with their responses to the past events (contrasted with no difference; i.e., 0). Although participants clearly relied on the past when evaluating future-oriented emotions (see past-to-future reasoning analysis above), it was not the case that they expected past emotions to simply re-instantiate. That is, future-oriented emotion judgments differed from emotion ratings for initial events ($|t|$s > 6.93, p < .001, ds > 0.56), recent events ($|t|$s > 2.08, ps < .039, ds > 0.17),[16] and the average of the past events ($|t|$s > 3.84, ps < .001, ds > 0.31).

[16]Except for PP trials, t[155] = 1.81, p = .073, d = 0.14.

We then explored which past event(s) adults prioritized when reasoning about uncertain future-oriented events. We calculated Cohen's ds and the associated confidence intervals for each difference score (larger effects show a greater difference between past and future emotion ratings). We judged effect sizes as different from one another when a given effect size fell outside of the confidence interval of another effect size. As a strict test of the recency bias in mixed-valence pasts (NP, PN), we tested whether future-oriented emotion ratings most closely aligned with emotion ratings for the recent past event (as opposed to how they felt in the initial episode or on average across the two events). Patterns indicated a recency bias for PN trials (Initial: d = 2.80, 95% confidence interval [CI] [2.45, 3.14]; Recent: d = 0.67, CI [0.50, 0.84]; Past Average: d = 1.75, CI [1.50, 2.00]). For NP trials, however, future-oriented emotions most closely resembled the past average (Initial: d = 1.44, CI [1.21, 1.66]; Recent: d = 0.82, CI [0.64, 1.00]; Past Average: d = 0.43, CI [0.26, 0.59]).

To clarify whether these patterns were simply driven by the valence of the recent past, we also examined consistent-past trials (NN, PP). Adults were equally like to rely on the recent past as on the past average emotion rating when the perpetrator only acted negatively (NN; Initial: d = 0.56, CI [0.39, 0.72]; Recent; d = 0.17, CI [0.01, 0.32]; Past Average: d = 0.31, CI [0.15, 0.47]). Adults showed a recency bias when the perpetrator only behaved positively (PP; Initial: d = 0.56, CI [0.39, 0.73]; Recent: d = 0.14, CI [-0.01, 0.30]; Past Average: d = 0.45, CI [0.28, 0.61]). Thus, the interaction between valence and sequence determines how adults draw from past emotional events to inform their future-oriented emotions.

Table 6. Chapter 2: Study 2 Means and Standard Deviations for the Difference Between Future Emotional Reactions and Emotional Reactions to Event 1, Event 2, and the Average of Event 1 and Event 2 by Social Group Membership

	LGBTQ-Latinx	Straight-Latinx	LGBTQ-White	Straight-White	All Participants
	Weighting the Past When Forecasting the Future				
PP Future – Initial	0.30 (0.53)	0.21 (0.74)	0.61 (0.66)	0.27 (0.45)	0.34 (0.61)
PP Future – Recent	0.06 (0.55)	0.04 (0.51)	0.10 (0.48)	0.07 (0.31)	0.07 (0.47)
PP Future – Past	0.18 (0.47)	0.13 (0.52)	0.35 (0.50)	0.17 (0.28)	0.21 (0.45)
NP Future – Initial	1.80 (1.15)	1.41 (1.07)	1.44 (1.23)	1.91 (1.12)	1.66 (1.15)
NP Future – Recent	-0.59 (0.77)	-0.99 (1.25)	-1.06 (1.26)	-0.95 (1.01)	-0.88 (1.08)
NP Future – Past	0.60 (0.75)	0.21 (0.98)	0.19 (1.02)	0.48 (0.88)	0.39 (0.91)
PN Future – Initial	-3.34 (1.39)	-3.64 (1.32)	-3.51 (1.16)	-3.54 (1.13)	-3.50 (1.25)
PN Future – Recent	0.59 (0.82)	0.36 (0.67)	0.43 (0.78)	0.56 (0.65)	0.49 (0.74)
PN Future – Past	-1.38 (1.00)	-1.64 (0.78)	-1.54 (0.86)	-1.49 (0.75)	-1.50 (0.86)
NN Future – Initial	-0.24 (0.87)	-0.46 (0.62)	-0.47 (0.76)	-0.48 (0.61)	-0.40 (0.73)
NN Future – Recent	0.17 (0.66)	0.00 (0.42)	0.00 (0.41)	0.13 (0.42)	0.08 (0.50)
NN Future – Past	-0.03 (0.69)	-0.23 (0.40)	-0.24 (0.50)	-0.17 (0.40)	-0.16 (0.52)

Note. PP = Positive then Positive; NP = Negative then Positive; PN = Positive then Negative; NN = Negative then Negative. "Initial" = emotion rating for first event in the sequence; "Recent" = emotion rating for second event in the sequence; "Past" = (emotion rating for initial event + emotion rating for recent event) / 2. Positive scores indicate that the future is more positive than the past; negative scores indicate that the future is more negative than the past; scores no different from 0 indicate that the past and present are equivalent.

General Discussion

Emerging adults who identified as LGBTQ-Latinx, Straight-Latinx, LGBTQ-White, or Straight-White judged that past emotional episodes shape affective reactions to events across time: Participants in all groups modified their emotional reactions to seemingly unambiguous positive and negative events based on the preceding event. This reliance on the initial past was stronger for negative compared to positive prior events. Although participants drew from the past when forecasting the future (more positive affective responses for PP > NP > PN > NN), future-oriented emotions differed from past emotional reactions, indicating that participants did not simply expect past emotions to reinstate. Signaling a recency bias, future-oriented emotions more closely aligned with emotion ratings from the recent versus initial past event. All findings were robust to sexual orientation and ethnic social group membership (the central results replicated in

all four groups). Below, we integrate findings from the new Emotions Across Time Task (EATT) with related research and consider directions for future investigations.

Judgments about Emotions Across Time

Adults consider information beyond the current event to infer their own and others' emotions, including beliefs, thoughts, and expectations (Atkinson et al., 2009; Lagattuta, 2014; Lagattuta et al., 2015; Lara et al., 2019; Shepperd & McNulty, 2002; Mrkva et al., 2019; Payir & Guttentag, 2016; Roese, 1997). We identified another factor that adults utilize when determining emotions: They expect the past to bias reactions to emotional events. In particular, adults judged that past positive events make subsequent positive events feel even *better* (PP) and past negative events make subsequent negative events feel even *worse* (NN). In addition to this augmenting effect, we documented a negativity bias when participants reasoned about events that differed in valence from the initial event to the subsequent (recent) event in sequence. Whereas initial (past) negative events greatly attenuated positive responses to recent (past) positive events (NP), initial (past) positive events had no influence on reactions to recent (past) negative events (PN). This contamination of negative experiences on subsequent positive events extends research showing that negative information is stickier and more difficult to overcome than positive information in impression formation, decision making, and attitude change (Bizer & Petty, 2005; Boydstun et al., 2018; Klein & O'Brien, 2016; Ledgerwood & Boydstun, 2014; O'Brien & Klein, 2017; Skowronski & Carlston, 1992; Sparks & Ledgerwood, 2018; Ferguson et al., 2019).

It is worth stressing that participants judged positive events as unambiguously positive and negative events as unambiguously negative when they were in isolation or occurred first in a sequence (i.e., across two studies they rated positive events as feeling medium good and negative events as feeling medium bad). These findings make their responses to a positive event that

followed a negative event particularly compelling. For NP pasts, the average emotion rating for the positive event was weaker than "a little good" with 42% of participants endorsing a negative emotional reaction to this positive event. In contrast, only 4% of participants rated the negative event positively for a PN trial. Adults may have viewed perpetrators' negative prior behaviors as more intentional and diagnostic of character than their more socially normative positive actions (Knobe, 2003; Skowronski & Carlston, 1989)—making negative events feel unequivocally negative, but positive events as more up to interpretation depending upon the preceding circumstances (Cone & Ferguson, 2015; Ferguson et al., 2019). Thus, when adults imagined experiencing the positive event in the NP trial, they may have questioned whether the positive event was still definitively rewarding. For example, if someone made fun of you, but then helped you, you may interpret the helping to indicate the person deems you incompetent. In contrast, if instead this person helped you and then later made fun of you, being ridiculed still feels bad.

Relying on the Past when Thinking about the Future

Research crossing a wide range of topics including economics, politics, business, and health indicate that people look to the past to forecast the future (Karinol & Ross, 1996; Malmendier & Nagel, 2011; Sönmez & Graefe, 1998; Suddendorf & Corballis, 2007; Ward et al., 2013). Not surprisingly, then, participants relied on prior emotional episodes when reasoning about their affective reactions to uncertain future-oriented events. Conceptually replicating Lagattuta and Sayfan (2013), participants judged that they would feel more intensely positive, think more optimistically, and make more confident approach decisions re-encountering the instigator of PP > NP > PN > NN pasts. The design of the current study afforded the opportunity to explore additional aspects of how adults integrate past events when thinking about the future. We tested whether future-oriented emotional reactions were simply a reinstatement of one or

both prior emotional reactions. This was not the case. Instead, while future-oriented emotion judgments shared similarity with past emotions, they were not equivalent (e.g., in PN trials, participants rated their future-oriented emotion more negatively than their response to the initial past event, more positively than their response to the recent past event, and more negatively than the average of the past two events). These data suggest that there is something unique about future forecasting that cannot be fully captured in past experiences. One candidate for the cause of these differences is that adults may try to account for the uncertainty of the future (e.g., Lagattuta & Sayfan, 2011). For example, even when the perpetrator previously behaved consistently negatively, some participants left open the possibility that the past does not constrain the future. Additional work using computational modeling to test more precisely how participants think about the future based on the past (e.g., including an indicator of uncertainty) would be informative (e.g., Ong et al., 2019).

Our results further reveal which past event participants prioritize when considering the future. Although adults weighted recent events more than initial events, the findings did not provide unwavering support for an exclusive recency heuristic. For NP trials, the average emotion rating of the two past events more closely approximated future-oriented emotion judgments than the recent event alone. Moreover, whereas in Lagattuta and Sayfan (2013) participants expected future-oriented emotions to align with the most recent past (e.g., feel positive after NP past, feel negative after PN past), adults in our study reported that they would feel closer to neutral seeing the NP perpetrator (but still feel bad seeing the PN perpetrator). In Lagattuta and Sayfan (2013), participants were told how the character felt after each past event and only predicted their future-oriented reactions. This may have suggested that the character had "moved on" and did not let the first episode permeate how they reacted to subsequent events.

The current paradigm provided a stricter test of the recency bias because participants rated emotions for each past event—permitting them to carry the initial emotional event through time. That they pushed forward negative initial events but not positive initial events speaks to the salience and temporal stickiness of negative emotional information (Baumeister et al., 2001; Lagattuta & Kramer, 2017; Ledgerwood & Boydstun, 2014; Vaish et al., 2008).

Social Group Membership and Emotion Cognition

Social group membership did not moderate judgments about emotional responses to events in isolation, events in sequence, or future-oriented events. Thus, in contrast to our hypotheses and prior research (Crocker et al., 1998; Inzlicht et al., 2009), we did not find any evidence that a general negativity bias was stronger in the marginalized social groups that we tested (i.e., LGBTQ-Latinx, LGBTQ-White, and Straight-Latinx people) than in the non-marginalized social group (i.e., Straight-White individuals). Given that past work demonstrates that members of marginalized groups experience more instances of discrimination (e.g., Bauer & Scheim, 2019; Paradies et al., 2015), the current findings suggest that, at the group level, these experiences do not necessarily shape people's perceptions of or expectations about interpersonal situations that are not explicitly related to their social group status. Importantly, then, these results argue against a "victimhood mentality" that is sometimes attributed to members of marginalized groups by people with privilege and power (Dwyer, 2014; Marshall, 2010; Talburt, 2004). Rather, individuals from marginalized social groups have emotional and social cognitive processes that are (at least) as nuanced and balanced as those from majority groups. Therefore, the tasks in the current research appear to be equally applicable across diverse social groups, speaking to their potential utility for other researchers interested in assessing emotion cognition.

Bridging to research on intercategorical intersectionality can inform extensions of the

EATT. For example, Scheim and Bauer (2019) found that LGBTQ people of color anticipated negative events more often than Straight-White participants after participants were instructed to think about their social identities (e.g., skin color, ancestry, gender, sexuality). In that research, however, participants were told to envision their identity as the cause of the negative event (e.g., *"Because of who I am, people might try to attack me physically*"). Thus, combining the EATT with the approach of Scheim and Bauer (2019) might elucidate the specific contexts where social group differences in negativity biases emerge. Recall that in the current work, participants in Study 1 reported demographic information (including sexual orientation and ethnic identity) immediately prior to answering questions about how they would emotionally react to events. Participants in Study 2 also knew that they were part of a larger longitudinal study aimed to measure psychosocial stressors related to social group identity. Despite this potential priming of social identities (DeMarree et al., 2005; Gaither et al., 2013), there were no group differences in responses to events in isolation or on the EATT. Thus, we anticipate that social identity priming needs to be robust and linked to the cause of an event for groups to differ in their responses.

Alternatively, it is possible that participants purposely reported the emotion, thought, and decision judgments that they thought conformed to some population-wide average, rather than to their personal experiences and perspectives. We find this interpretation implausible for two reasons. First, it would have required participants to estimate accurately how people from other social groups would respond to each event, and then systematically adjust their ratings across multiple trials to fit those anticipated answers; an incredible perspective-taking feat. Moreover, participants in Study 1 and Study 2 gave equivalent responses despite only the latter group being part of the longitudinal study. Although we did not document group-level differences, there was clear *within* group variability on the EATT, especially for NP trials (e.g., some individuals

anticipated feeling negatively during the recent positive event and/or the future-oriented emotion whereas others provided positive emotion ratings; see *Figures 2* and *3*). To understand these individual differences, in future work we will analyze relations among participants' EATT responses, their personal experiences with discrimination, and mental health (Bauer & Scheim, 2019; Parra & Hastings, 2020; Scheim & Bauer, 2019).

Limitations and Future Directions

When thinking about how the past influences reactions to future-oriented events, adults attributed similar emotional responses to themselves (current research) and to others (Lagattuta & Sayfan, 2013). This corresponds with research suggesting that biases in affective forecasting influence people's beliefs about self and other emotions in similar ways (O'Brien et al., 2018). Other work, however, suggests that thinking about one's own and others' reactions can lead to differing judgments (Ong et al., 2018). Research systematically examining when emotional perspective taking in the first- versus third-person converges or diverges would be interesting. More broadly, the importance of replication should not be overlooked. With the rising concern that several key findings in psychological research are not replicable (Open Science Collaboration, 2015), it is noteworthy that this is the second successful conceptual replication of how adults integrate past emotional episodes when reasoning about future-oriented affective states (i.e., PP > NP > PN > NN pasts; see also Lagattuta & Kramer, 2019). Of course, independent replications outside of our lab would be critical. The creation of this online version of EATT will aid in this goal, as it will be easier to share these materials with other researchers. Indeed, the EATT could be modified to address additional questions about complex emotion cognition. For example, it would be informative to test whether varying the temporal spacing of the initial and recent events (e.g., days, weeks, years) impacts judgments, and to examine how

participants respond to longer sequences of past events. As well, examining whether responses to the EATT vary depending upon a participant's relationship to the perpetrator (e.g., stranger, parent, friend) would be an important focus for future studies.

In Lagattuta and Sayfan (2013), children as young 4 to 5 years recognized the biasing impact of past emotional events on how a person would feel, think, and make decisions seeing that perpetrator again. It remains unclear, however, at what age children appreciate that past experiences can also influence emotional reactions to seemingly unequivocal negative and positive events. Moreover, because young children (especially those under 6 years of age), tend to have high confidence when predicting the future (Lagattuta & Sayfan, 2011), it is possible that the correspondence between their future-oriented reactions and past emotional reactions would be more similar than they are for adults because children leave less room for uncertainty. It is further unknown whether young children would reason about story characters the same way as they reason about themselves (as did the adults in the current research). Future studies are needed that incorporate a developmental perspective by testing children's and adults' responses to the EATT (or more simplified versions of the EATT) across a wide age range.

Our goal was to assess adults' beliefs about emotions across time, but these intuitions may not correspond to the ground truth of their actual emotional reactions. Adults often struggle to accurately infer how they will feel at later points in time: They overestimate the type, intensity, and duration of their affective reactions (Kramer et al., 2017; Kramer & Lagattuta, 2018; Meyvis et al. 2010; Wilson & Gilbert, 2003). Thus, it is possible that the effect of past experiences on subsequent reactions to negative, positive, and uncertain future events would be weaker or more nuanced than participants reported. Moreover, we made the past events salient by asking participants to rate their emotional reactions after each event and reminding them

about the two events before they provided their future-oriented responses. This scaffolded context may have artificially highlighted the valence consistency (or inconsistency) of the sequenced events. It would be interesting to explore the extent to which people spontaneously reflect on their emotions after an event as well as their accuracy at tracking the sequence of negative and positive events in their everyday lives. Indeed, although we found no social group differences in adults' judgments about emotions across time, this does not preclude the possibility that differences could emerge in their first-person experiences in the lab or in real life.

Conclusions

Adults share strong consensus about emotional events in isolation: They judge that negative events feel bad and positive events feel good. Still, they also appreciate that emotional reactions are not always linearly and directly derived from the present circumstances. Instead, they reason that emotions carry forward in time to influence how they later respond to seemingly unambiguous positive and negative events as well as to uncertain future-oriented events, with prior negative episodes being especially sticky. Although there was variability in the extent to which adults expected initial events to contaminate subsequent reactions, emotional ratings to events in isolation and on the EATT replicated at the group level in LGBTQ-Latinx, Straight-Latinx, LGBTQ-White, and Straight-White emerging adults. The current methods and data lay a foundation for new empirical questions for the study of emotion. Focally, they indicate that affective scientists need to attend not only to the "objective" features of emotional events when designing procedures, but also consider the complex reality that individuals rarely experience events in isolation; rather, we build our emotional responses across time.

References

Asch, S.E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*, 258-290. doi: 10.1037/h0055756

Atkinson, L., Bell, D., & Feeney, A. (2009). The relationship between counterfactual thinking and emotional reactions to event outcomes: Does one account fit all? *Psychonomic Bulletin & Review, 16*, 724-728. doi: 10.3758/PBR.16.4.72

Bamford, C., & Lagattuta, K.H. (2012). Looking on the bright side: Children's knowledge about the benefits of positive versus negative thinking. *Child Development, 83*, 667-682. doi: 10.1111/j.1467-8624.2011.01706.x

Bauer, G. R., & Scheim, A. I. (2019). Methods for analytic intercategorical intersectionality in quantitative research: Discrimination as a mediator of health inequalities. *Social Science & Medicine, 226*, 236-245. doi: 10.1016/j.socscimed.2018.12.015

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., & Vohs, K.D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323-370. doi: 10.1037//1089-2680.54.323

Bizer, G.Y., & Petty, R.E. (2005). How we conceptualize our attitudes matters: The effects of valence framing on the resistance of political attitudes. *Political Psychology, 26*, 553-568. doi: 10.1111/j.1467-9221.2005.00431.x

Boydstun, A.E., Ledgerwood, A., & Sparks, J. (2019). A negativity bias in reframing shapes political preferences even in partisan contexts. *Social Psychology and Personality Science, 10*, 53-61. doi: 10.1177/1948550617733520

Brondolo, E., Kelly, K.P., Coakley, V., Gordon, T., Thompson, S., Levy, E.,…Contrada, R.J. (2005). The perceived ethnic discrimination questionnaire: Development and preliminary validation of a community version. *Journal of Applied Social Psychology, 35*, 335-365.

doi: 10.1111/j.1559-1816.2005.tb02124.x

Cole, E. R. (2009). Intersectionality and research in psychology. *American Psychologist, 64(*3), 170-180. doi: 10.1037/a0014564

Collins, P. H. (2002). Black feminist thought: Knowledge, consciousness, and the politics of empowerment. Routledge, New York.

Cone, J., & Ferguson, M.J. (2015). He did what? The role of diagnosticity in revising implicit evaluations. *Journal of Personality and Social Psychology, 108,* 37-57. doi: 10.1037/pspa0000014

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. University of Chicago Legal Forum, 139-167.

Crocker, J., Major, B., & Steele, C. (1998). Social stigma. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 504-553). New York, NY, US: McGraw-Hill.

DeBlaere, C., Brewster, M. E., Sarkees, A., & Moradi, B. (2010). Conducting research with LGB people of color: Methodological challenges and strategies. *The Counseling Psychologist, 38*(3), 331-362. doi: 10.1177/0011000009335257

DeMarree, K.G., Wheeler, S.C., & Petty, R.E. (2005). Priming a new identity: Self-monitoring moderates the effects of nonsense primes on self-judgments and behavior. *Journal of Personality and Social Psychology, 89*, 657-671. doi: 10.1037/0022-3514.89.5.657

Dwyer, A. (2014). "We're not like these weird feather boa-covered AIDS-spreading monsters": How LGBT young people and service providers think riskiness informs LGBT youth–police interactions. *Critical Criminology, 22*(1), 65-79. doi: 10.1007/s10612-013-9226-z

Ferguson, M.J., Mann, T.C., Cone, J., & Shen, X. (2019). When and how implicit impressions can be updated. *Current Directions in Psychological Science, 28,* 331-336. doi: 10.1177/0963721419835206

Gaither, S.E., Sommers, S.R., & Ambady, N. (2013). When the half affects the whole: Priming identity for biracial individuals in social interactions. *Journal of Experimental Social Psychology, 49,* 368-371. doi: 10.1016/j.jesp.2012.12.012

Garnets, L. D. (2002). Sexual orientations in perspective. *Cultural Diversity and Ethnic Minority Psychology, 8,* 115-129. doi: 10.1037/1099-9809.8.2.115

Goffman E. (1963) *Stigma*. Prentice Hall, Englewood Cliffs, NJ.

Gawronski, B., & Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692-731. doi: 10.1037/0033-2909.132.5.692

Klein, N., & O'Brien, E. (2016). The tipping point of moral change: When do good and bad acts make good and bad actors. *Social Cognition, 34,* 149-166. doi: 10.1521/soco/2016.34.2.149

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63,* 190-194. doi: 10.1111/1467-8284.00419

Kramer, H.J., Goldfarb, D., Tashjian, S.M., & Lagattuta, K.H. (2017). "These pretzels are making me thirsty": Older children and adults struggle with induced episodic foresight. *Child Development, 88,* 1554-1562. doi: 10.1111/cdev.12700

Kramer, H.J., & Lagattuta, K.H. (2018). Affective forecasting. In M. Bornstein, M.E. Arterberry, K.L. Fingerman, & J.E. Lansford (Eds.), The Sage encyclopedia of lifespan human development (71-72). Sage.

Kramer, H. J., & Lagattuta, K. H. (in press). Developmental changes in emotion understanding during middle childhood. To appear in D. D. Dukes, E. Walle, & A. C. Samson, (Eds.), *Oxford handbook on emotional development*, Oxford University Press.

Institute of Medicine (IOM). (2011). *The health of lesbian, gay, bisexual, and transgender people: Building a foundation for better understanding*. Washington, DC: The National Academies Press.

Inzlicht, M., Aronson, J., & Mendoza-Denton, R. (2009). On being the target of prejudice: Educational implications. In F. Butera & J. Levine (Eds.) *Coping with minority status: Responses to exclusion and inclusion* (pp. 13–36). Cambridge, UK: Cambridge University Press.

Jones, E. E., Farina, A., Hastorf, A. H., Markus, H., Miller, D. T., & Scott, R. A. (1984). *Social stigma: the psychology of marked relationships*. New York: W.H. Freeman & Company.

Lagattuta, K.H. (2007). Thinking about the future because of the past: Young children's knowledge about the causes of worry and preventative decisions. *Child Development, 78*, 1492-1509. doi: 10.1111/j.1467-8624.2007.01079.x

Lagattuta, K.H., & Kramer, H.J. (2017). Try to look on the bright side: Children and adults can (sometimes) override their tendency to prioritize negative faces. *Journal of Experimental Psychology: General, 146*, 89-101. doi: 10.1037/xge0000247

Lagattuta, K.H. & Kramer, H.J. (2019). Advanced emotion understanding: Children's and adults' knowledge that minds generalize from prior emotional events. *Emotion, advanced online publication.* doi: 10.1037/emo0000694

Lagattuta, K.H., Kramer, H.J., Kennedy, K., Hjortsvang, K., Goldfarb, D., & Tashjian, S.M. (2015). Beyond Sally's missing marble: Further development in children's understanding

of mind and emotion during middle childhood. *Advances in Child Development and Behavior, 48,* 185-217.

Lagattuta, K.H., & Sayfan, L. (2011). Developmental changes in children's understanding of future likelihood and uncertainty. *Cognitive Development, 26,* 315-330. doi: 10.1016/j.cogdev.2011.09.004

Lagattuta, K.H., & Sayfan, L. (2013). Not all past events are equal: Biased attention and emerging heuristics in children's past-to-future forecasting. *Child Development, 84,* 2094-2111. doi: 10.1111/cdev.12082

Lagattuta, K.H., Tashjian, S.M., Kramer, H.J. (2018). Does the past shape anticipation for the future? Contributions of age and executive function to advanced theory of mind. *Zeitschrift Für Psychologie, 226,* 122-133. doi: 10.1027/2151-2604/a000328

Lagattuta, K.H., & Wellman, H.M. (2001). Thinking about the past: Early knowledge about links between prior experience, thinking, and emotion. *Child Development, 72,* 82-102. doi: 10.1111/1467-8624.00267

Lagattuta, K. H., & Wellman, H. M. (2002). Differences in early parent-child conversations about negative versus positive emotions: Implications for the development of psychological understanding. *Developmental Psychology, 38,* 564-580. doi:10.1037/0012-1649.38.564

Lagattuta, K. H., Wellman, H. M., & Flavell, J. H. (1997). Preschoolers' understanding of the link between thinking and feeling: Cognitive cuing and emotional change. *Child Development, 68,* 1081-1104. doi: 10.2307/1132293

Lara, K.H., Lagattuta, K.H., & Kramer, H.J. (2019). Is there a downside to anticipating the upside? Children's and adults' reasoning about how prior expectations shape future

emotions. *Child Development, 90,* 1170-1184. doi: 10.1111/cdev.12994

Ledgerwood, A., & Boydstun, A.E. (2014). Sticky prospects: Loss frames are cognitive stickier than gain frames. *Journal of Experimental Psychology: General, 143*, 376-385. doi: 10.1037/a0032310

LoBue, V., & DeLoache, J.S. (2008). Detecting the snake in the grass: Attention to fear-relevant stimuli by adults and young children. *Psychological Science, 19,* 284-289. doi: 10.1111/j.1467-9280.2008.02081.x

Marshall, D. (2010). Popular culture, the 'victim' trope and queer youth analytics. *International Journal of Qualitative Studies in Education, 23*, 65-85. doi: 10.1080/09518390903447176

McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of Women in Culture and Society 30*, 1771–1800. doi: 10.1086/426800

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychological Bulletin*, *129*, 674-697. doi: 10.1037/0033-2909.129.5.674

Meyer, I. H. (1995). Minority stress and mental health in gay men. Journal of health and social behavior, 38-56. doi: 10.2307/2137286

Meyvis, T., Ratner, R.K., & Levav, J. (2010). Why don't we learn to accurately forecasting feelings? How misremembering blinds us to past forecasting errors. *Journal of Experimental Psychology: General, 139*, 579-589. doi: 10.1037/a0020285

Mrkva, K., Westfall, J., & Van Boven, L. (2019). Attention drives emotion: Voluntary visual attention increases perceived emotional intensity. *Psychological Science, 30,* 942-954. doi: 10.1177/0956797619844231

O'Brien, E., & Klein, N. (2017). The tipping point of perceived change: Asymmetric thresholds

in diagnosing improvement versus decline. *Journal of Personality and Social Psychology 112*, 161-185. Doi: 10.1037/pspa0000070

O'Brien, E., Kristal, A.C., Ellsworth, P.C., & Schwarz, N. (2018). (Mis)imagining the good life and the bad life: Envy and pity as a function of the focusing illusion. *Journal of Experimental Social Psychology, 75*, 41-53. doi: 10.1016/j.jesp.2017.10.002

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*, 381-396. doi: 10.1037/0022-3514.80.3.381

Ong, D.C., Goodman, N.D. & Zaki (2018). Happier than thou? A self-enhancement bias in emotion attribution. *Emotion, 18*, 116-126. doi: 10.1037/emo0000309

Ong, D.C., Zaki, J., & Goodman, N.D. (2019). Computational models of emotion inference in theory of mind: A review and roadmap. *Topics in Cognitive Science, 11*, 338-357. doi: 10.1111/tops.12371

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*. doi: 10.1126/science.aac4716

Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., ... & Gee, G. (2015). Racism as a determinant of health: a systematic review and meta-analysis. *PloS One, 10*, e0138511.10.1371/journal.pone.0138511

Parent, M. C., DeBlaere, C., & Moradi, B. (2013). Approaches to research on intersectionality: Perspectives on gender, LGBT, and racial/ethnic identities. *Sex Roles*, *68*, 639-645. doi: 10.1007/s11199-013-0283-210.1002/cad.20244

Parra, L. A., & Hastings, P. D. (2020). Challenges to Identity Integration Indirectly Link Experiences of Heterosexist and Racist Discrimination to Lower Waking Salivary

Cortisol in Sexually Diverse Latinx Emerging Adults. *Frontiers in Psychology, 11*, 228. doi: 10.3389/fpsyg.2020.00228

Payir, A., & Guttentag, R. (2016). "It could have been worse": Developmental change in the use of counterfactual consoling strategy. *Journal of Experimental Child Psychology, 148*, 119-130. doi: 10.1016/j.jecp.2016.04.001

Qualtrics (2019). Qualtrics Survey Software. Provo, UT.

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.

Roese, N.J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*, 133-148. doi: 10.1037/0033-2909-.121.1.133

Rosario, M., Schrimshaw, E.W., Hunter, J., & Gwadz, M. (2002). Gay-related stress and emotional distress among gay, lesbian, and bisexual youths: A longitudinal examination. *Journal of Consulting and Clinical Psychology, 70*, 967-975. doi: 10.1037/0022-006X.70.4.967

Scharrón-del Río, M. R., & Aja, A. A. (2015). The case FOR 'Latinx': Why intersectionality is not a choice. *Latino Rebels,* 5. Retrieved September 4, 2019, from http:// www.latinorebels.com/2015/12/05/the-case-for-latinx-why-intersectionality-is-not-a-choice/.

Scheim, A. I., & Bauer, G. R. (2019). The Intersectional Discrimination Index: development and validation of measures of self-reported enacted and anticipated discrimination for intercategorical analysis. *Social Science & Medicine, 226*, 225-235. doi: 10.1016/j.socscimed.2018.12.016

Shepperd, J.A., & McNulty, J.K. (2002). The affective consequences of expected and unexpected outcomes. *Psychological Science, 13*, 85-88. doi: 10.1111/1467-9280.00146

Skowronski, J.J., & Carlson, D.E. (1989). Negativity and extremity biases in impression formation. A review of explanations. *Psychological Bulletin, 105*, 131-142. doi:

10.1037/0033-2909.105.1.131

Skowronski, J.J. & Carlston, D.E. (1992). Caught in the act: When impressions based on highly diagnostic behaviours are resistant to contradiction. *European Journal of Social Psychology, 22,* 435-452. doi: 10.1002/ejsp.2420220503

Sparks, J., & Ledgerwood, A. (2019). Age attenuates the negativity bias in reframing effects. *Personality and social Psychology Bulletin, 45*, 1042-1056. doi: 10.1177/0146167218804526

Talburt, S. (2004). Constructions of LGBT youth: Opening up subject positions. *Theory into Practice, 43*, 116-121. doi: 10.1207/s15430421tip4302_4

Uleman, J.S., & Kressel, L.M. (2013). A brief history of theory and research on impression formation. In D.E. Carlston (Ed.). *Oxford handbook of social cognition* (pp. 55-73). New York: Oxford University Press.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*, 383-403. doi: 10.1037/0033-2909.134.3.383

Ward, M.D. Metternich, N.W., Dorff, C.L., Gallop, M., Hollenbach, F.M., Schultz, A., & Weschle, S. (2013). Learning from the past and stepping into the future: Toward a new generation of conflict prediction. *International Studies Review, 15*, 473-490. doi: 10.1111/misr.12072

Wilson, T.D., & Gilbert, D.T. (2003). Affective forecasting. In M.P. Zanna (Ed.), *Advances in experimental social psychology,* (pp. 345-411). Academic Press.

Chapter 3

Dichotomous Thinking about Social Groups:

Learning about One Group Can Activate Opposite Beliefs about Another Group

Abstract

Across three studies (N = 607), we examined people's use of a *dichotomizing heuristic*—the inference that characteristics belonging to one group do not apply to another group—when making judgments about novel social groups. Participants learned information about one group (e.g., "Zuttles like apples"), and then made inferences about another group (e.g., "Do Twiggums like apples or hate apples?"). Study 1 acted as a proof of concept: Eight-year-olds and adults (but not 5-year-olds) assumed that the two groups would have opposite characteristics. Learning about the group as a generic whole versus as specific individuals boosted the use of the heuristic. Study 2 and Study 3 (sample sizes, methods, and analyses pre-registered), examined whether the presence or absence of several factors affected the activation and scope of the dichotomizing heuristic in adults. Whereas learning about or treating the groups as separate was necessary for activating dichotomous thinking, intergroup conflict and featuring only two (versus many) groups was not required. Moreover, the heuristic occurred when participants made both binary and scaled decisions. Once triggered, adults applied this cognitive shortcut widely—not only to benign (e.g., liking apples) and novel characteristics (e.g., liking modies), but also to evaluative traits signaling the morals or virtues of a social group (e.g., meanness or intelligence). Adults did not, however, extend the heuristic to the edges of improbability: They failed to dichotomize when doing so would attribute highly unusual preferences (e.g., disliking having fun). Taken together, these studies indicate the presence of a dichotomizing heuristic with broad implications for how people make social group inferences.

Keywords: social cognition, categories, generic language, development, heuristics

Dichotomous Thinking about Social Groups:

Learning about One Group Can Activate Opposite Beliefs about Another Group

People often expect social groups to differ—for example, they believe that members from one group will hold values, beliefs, preferences, skills, and traits that are distinct from members of another group. Such inferences about between-group differences, whether accurate or not, have been shaped over time through systemic influences like social hierarchies, stereotypes, and prejudice, as well as by directly experienced or observed social interactions and information from others (Richter et al., 2016; Schwarz & Bless, 2007; Tajfel, 1982). In the current study, we examined the inferences individuals make when first learning about a novel social group. In the absence of complex historical and phenomenological evidence, will people assume that different social groups hold opposite characteristics?[17]

Across several domains, people often use heuristics, or cognitive shortcuts, to make judgments because they are quicker than more thorough, careful, and deliberate thinking (Kahneman, 2011; Sherman & Corty, 1984; Sunstein, 2005; Tversky & Kahneman, 1973). Heuristics play a vital role in reducing uncertainty, decreasing cognitive load, and increasing processing efficiency. They can, however, also lead to biased assumptions about social groups (e.g., when forming an impression of a new person from a known social group, people often use stereotypes of that social group as an anchor; when making adjustments away from this anchor they do not go far enough; FeldmanHall & Shenhav, 2019). Here, we investigate whether people apply a *dichotomizing heuristic*—what is true of one group is not true of another group—when

[17]This paper is an accepted manuscript. Citation: Kramer, H. J., Goldfarb, D., Tashjian, S. M., & Lagattuta, K. H. (in press). Dichotomous thinking about social groups: Learning about one group can activate opposite beliefs about another group. To appear in *Cognitive Psychology*.

forming impressions of novel social groups for which they have limited information.[18] To test this potential heuristic, participants learned about the preferences, abilities, and traits of one group (*characterized group*). Participants then made inferences about an *uncharacterized group's* preferences, abilities, and traits (no information provided). Across three studies, we examined what triggers the dichotomizing heuristic as well as its scope.

Humans assume that individuals who are members of the same group share common features (Gelman, 2003). By three to four years of age, children use category labels to make inferences about the biological and psychological properties of individuals; for example, they infer that two creatures called "fish" will both breathe under water or that two kids categorized as "shy" will both like the same game (Gelman & Markman, 1986; Heyman & Gelman, 2000a, 2000b). Adults also give particular weight to category labels as compared with other features that may signal group membership when drawing inferences (Yamauchi & Markman, 2000). Moreover, preschoolers, school-aged children, and adults tend to extend information learned about the characteristics of one category member to all category members. For example, they predict that individuals from the same social group (e.g., gender or race) will share similar traits, preferences, beliefs, and behaviors (Diesendruck et al., 2015; Kalish, 2012; Krueger & Clement, 1994; Lagattuta & Kramer, 2021; McGarty & Penny, 1988; McGarty & Turner, 1992; Shutts et al., 2013), with adults overestimating stereotypic traits among group members (e.g., all women are caring; Allport, 1954; Devine, 1989; Park & Rothbart, 1982). Thus, when individuals are members of the same category or group, people across a wide age range assume homogeneity: They infer that learning about the characteristics of one member provides valuable evidence about all members and that category labels can be informative about group characteristics.

[18]In labeling this a heuristic, we are not arguing that this inference is de facto irrational or rational, only that it is a rule-of-thumb approach that enables decisions under uncertainty with minimal mental effort.

Categories, however, not only prime beliefs about within-group cohesiveness, but can also exaggerate perceptions of between-group differences, sometimes called *category accentuation* (Eiser, 1971; Tajfel & Wilkes, 1963). For example, 4-year-olds who first sorted faces into two separate categories (mean vs. nice) versus on a scaled continuum (very mean to very nice) more often assumed greater differences in behaviors, preferences, and intentions of meaner versus nicer looking individuals (Master et al., 2012). Moreover, adults evaluate a group with neutral traits more favorably when they simultaneously learn about another group with negative traits compared to when they simultaneously learn about another group with positive traits (Krueger & Rothbart, 1990). Intriguingly, the presence of real-world group differences is not necessary for people to assume dissimilarities; there can be *illusory correlations* (Hamilton & Gifford, 1976). For example, children and adults tend to associate majority groups with more frequent behaviors (e.g., positive behaviors) and minority groups with less frequent behaviors (e.g., negative behaviors), even when the proportions of these actions are equivalent across groups (Johnston & Jacobs, 2003; Lawson & Bower, 2014; Sherman et al., 2009). Thus, when humans learn about multiple categories or groups, they tend to attribute greater between-group differences than warranted by observable or verifiable evidence.

Here, we extend inquiry into people's beliefs about social groups as categories. As reviewed, the typical paradigm has involved (1) telling participants characteristics of one group member and seeing if they extend those characteristics to all group members, or (2) providing information about two separate groups and testing if participants inflate between-group differences. But, what happens if participants only learn the characteristics of *one* group—what assumptions do they make about members of a separate, uncharacterized group? In this uncertain context where no details are provided about the uncharacterized group, do people

default to a dichotomizing heuristic—assume that what is true of one group must not be true of the other group? That is, with no descriptions of the uncharacterized group to draw from, will people judge that the two groups have opposite characteristics (evidence of a dichotomizing heuristic), assume the two groups share the same characteristics, or just guess randomly?

Study 1 served as a proof of concept: Our goal was to document whether people assume that they have acquired information about two groups—a characterized group and an uncharacterized group—when they have only learned about one. In this initial study, we further tested whether the language used to discuss the characterized group influences application of the dichotomizing heuristic. That is, we assessed whether use of the dichotomizing heuristic is more pronounced when information about the social category is presented generically (i.e., describing the characterized group as a whole; e.g., "Zuttles like pears") rather than specifically (i.e., describing individuals within a group; e.g., "This Zuttle, X, likes pears"). Generic language encourages children and adults to assume greater homogeneity within categories (Chambers et al., 2008; Gelman et al., 2002) and to create stricter boundaries between categories (Gelman et al., 2010; Goldfarb et al., 2017). Thus, we tested whether participants who learn about the characterized group as a generic whole (versus as specific individuals) more often exhibit the dichotomizing heuristic (assume that the uncharacterized group possesses opposite traits).

A further goal of Study 1 was to identify age-related differences in the application of the dichotomizing heuristic. Extant research suggests that this cognitive shortcut, if present, would develop with age as individuals continue to explore and form models of the social world. Preschoolers, compared to older children and adults, require more generic input before they generalize group features across all category members (Cimpian & Scott, 2012), and they form less rigid category boundaries between separate social groups (Shutts, 2015). Reliance on

heuristics also strengthens not only from early to middle childhood but continues to increase from childhood to adulthood (Jacobs & Potenza, 1991; Lagattuta & Kramer, 2021; Lagattuta & Sayfan, 2013; Lagattuta et al., 2018; Reyna & Ellis, 1994). Based on this combined evidence, we reasoned that young children would less often use a dichotomizing heuristic to make inferences about an uncharacterized group. Therefore, we tested whether older participants would more often endorse inter-group dichotomies than younger children.

Recent evidence to support this new research direction in social group reasoning comes from Moty and Rhodes (2021).[19] Children as young as 4.5 years, and increasing with age (i.e., between 4.5 and 7 years and between 7 and adulthood), judged that characteristics of one group (e.g., skill at baking) do not apply to a different social group. Moreover, learning about groups in generic (e.g., Zarpies are good at baking pizza) versus specific language (e.g., This Zarpie is good at baking pizza) magnified this effect. In these studies, however, the experimenter explicitly told participants that the two groups—Zarpies and Gorps—were "two kinds of people" and they pictured them wearing distinctive clothes (e.g., all green for Zarpies; all yellow for Gorps). Starting from the salient premise that the two groups had different origins and different clothing preferences could have biased participants to expect dissimilarity across all characteristics, facilitating dichotomizing at an early age.

In Studies 1 and 2, we provided a more stringent test by examining whether participants also exhibit the dichotomizing heuristic when the two social groups share a common biological and social origin (e.g., similar to how non-indigenous American colonists separated from the British) and cannot be distinguished visually; thus, making it equivalently plausible to assume

[19]For transparency, although we include Moty and Rhodes (2021) in the Introduction, the methods and hypotheses for the studies reported here were not created to extend that study. We completed all analyses for Study 1 prior to Moty and Rhodes' preregistration. Pilot Study 2 and Study 2 were preregistered and completed prior to Moty and Rhodes (2021) being accepted for publication. Moty and Rhodes (2021) also do not use the term "dichotomizing heuristic" when discussing their findings.

either between-group similarities or differences. In Study 2, we also manipulated exposure to a categorization (sorting) task to test whether thinking about the groups as separate entities activates the dichotomizing heuristic. In that study, we further investigated whether intergroup conflict or beneficence shaped the dichotomizing heuristic. In Study 3, we removed the shared biological and social origin of Zuttles and Twiggums, but we provided no distinguishing information beyond the different category labels. In this study, we again tested whether experience separating individuals into different groups (the categorization task) is necessary for activating the heuristic if groups were initially described as distinct. In addition, we measured whether the dichotomizing heuristic was bolstered by learning about the uncharacterized group in the presence of only one other group (i.e., setting up a binary contrast) versus in the presence of many groups (i.e., no clear contrast between one group and another). We also changed our dependent variable from binary to scaled to test the robustness of the effect. Moreover, we assessed the scope and potential limits of the dichotomizing heuristic by analyzing whether participants apply it to benign (e.g., liking apples), novel (e.g., like modies), evaluative (e.g., being smart), and universal characteristics (e.g., disliking getting hurt).

Study 1

Study 1 builds upon the novel-group paradigm utilized by Goldfarb et al. (2017; for further examples of novel-group methods, see Killen, 2007; Roberts et al., 2018). In Goldfarb et al. (2017), 5-year-olds, 8-year-olds, and adults learned about a fictitious social group of human-like creatures, Twiggums, who lived in a valley. Some Twiggums left the valley to live in the mountains. There, they formed a new group and renamed themselves Zuttles, a social group whose members still looked indistinguishable from Twiggums. Next, participants learned about Zuttle preferences and abilities either in generic language (e.g., “Zuttles like pears”) or in

specific language (e.g., “This Zuttle, Dax, likes pears”). Participants then completed a *social categorization task* where they saw a creature paired with three characteristics. Children and adults decided whether that individual was a Zuttle or a Twiggum by making characteristic-to-category judgments. They received no feedback about the correctness of their choice because there was no objective truth given that participants had no information about Twiggums.

In one version of the social categorization task (neutral stance), participants sorted creatures without consequence: Nothing happened to those identified as Zuttles. In the other condition (negative stance), a law banned creatures identified as Zuttles from the valley, forced them to be jailed for 100 days, and then required their deportation. While Goldfarb et al. (2017) presented the findings from the categorization task, Study 1 of the current manuscript focuses on a separate measure that participants completed after the social categorization task. This new *uncharacterized group task* asked children and adults to judge whether they thought Twiggums (uncharacterized group) differed from or aligned with Zuttles (characterized group) in their preferences and abilities; that is, they made category-to-characteristic judgments. Of central interest was whether participants would use a dichotomizing heuristic—what is true of one group is not true of the other group—when making inferences about the uncharacterized group. Our paradigm differed from Moty and Rhodes (2021) in that we neither described Zuttles and Twiggums as “two kinds of people” nor showed them as differing in appearance. Rather, to lessen the demand to assume group differences, Zuttles and Twiggums (1) had a common biological and social origin; and (2) looked identical. In addition to these shared features, we also included factors which could suggest that the groups might differ, including elective migration and punitive laws: (1) some Twiggums decided to move to the mountains and call themselves Zuttles; and (2) the Twiggums then made a law banning Zuttles from the valley.

Method

Participants

Participants included 181 children and adults divided into three age groups: 66 5-year-olds (*M* = 5.57 years, *SD* = 0.76 years, range = 4.20 years to 6.99 years; 34 male, 32 female), 62 8-year-olds (*M* = 8.40 years, *SD* = 0.84 years, range = 7.06 years to 10.48 years; 29 male, 33 female), and 53 adults (*M* = 22.61 years, *SD* = 7.47 years, range = 18.17 years to 60.15 years; 24 male, 29 female). The sample was 14% Asian, 2% Black or African American, 10% Hispanic/Latino, 58% White, and 16% "Other" race or ethnicity.[20] Included participants were fluent in comprehending and speaking English and were typically developing (parent-reported for child participants, self-reported for adults). Children were recruited from a list of previous participants, referral, recruitment at local farmers' markets and schools, listserv emails, and fliers. Adults were recruited from an undergraduate psychology database. Data were collected from October 2013 to October 2014. Participants were the same as those included in Goldfarb et al. (2017) minus nine adults who did not complete the uncharacterized group task. This study was approved by the Internal Review Board at [BLINDED]: Protocol Number: 448441.

Measures and Procedures

This study followed a 3 (age: 5-year-olds, 8-year-olds, adults) x 2 (language: generic, specific) design. Age and language were between-subjects factors.

Narrative and Social Categorization Task. Participants learned from a pre-recorded video (i.e., pictures on a computer screen accompanied by audio) that a group of human-like creatures, Twiggums, lived in a valley. Some of the Twiggums left the valley to live in the mountains. Once there, this group of Twiggums re-named themselves Zuttles. Participants then

[20]Race and ethnicity data were missing from one 8-year-old participant.

learned about Zuttles' preferences and abilities in either generic (e.g., "Zuttles like pears.") or specific language (e.g., "This Zuttle, Dax, likes pears."). All featured Zuttle characteristics were benign (i.e., did not indicate any deviance or threat of harm) and non-distinctive (i.e., could be widely shared, even with the participants themselves).

Next, participants completed the social categorization task that required them to determine whether a series of creatures were Zuttles (characterized group) or Twiggums (uncharacterized group) based on three pieces of information about their characteristics (only a subset matched what they learned about Zuttles). In the negative stance context, participants were told that the Twiggums had created a law forbidding Zuttles from living in the valley. Thus, those identified as Zuttles would be sent to jail for 100 days and then deported. In the neutral stance context, there were no consequences for creatures identified as Zuttles. Context order was counterbalanced. To control for differences in responding due to variation in working memory skills, all participants were given a memory-cue card that displayed the characteristics they had learned about Zuttles. Following research with young children (e.g., Lagattuta, 2008; Lara et al., 2019), we used crossed-out images to signify that a Zuttle/Zuttles hates something or is/are bad at something. For more details on the social categorization task, see Goldfarb et al. (2017).

Uncharacterized Group Task. After the social categorization task, children and adults predicted Twiggums' (uncharacterized group) preferences and abilities. We introduced the task as, "While we talked a lot about Zuttles, I am going to ask you some questions about what you think Twiggums might be like. Just make your best guess. There is no right answer. This is based on what you think." We then asked participants about eight of the characteristics (e.g., "Do you think Twiggums like pears or hate pears?"). The memory-cue sheet remained visible throughout.

We coded each response as *dichotomized* (1; participant said that Twiggums had *the*

opposite preference or ability from Zuttles) or *aligned* (0; participant said that Twiggums had *the same* preference or ability as Zuttles). The eight characteristics showed good internal consistency (Cronbach's alpha = .87, 95% confidence interval [CI] [.84, .90]; 5-year-olds: Cronbach's alpha = .80, 95% CI [.73, .87]; 8-year-olds: Cronbach's alpha = .89, 95% CI [.85, .93]; adults: Cronbach's alpha = .90, 95% CI [.86, .94]). We calculated the proportion of trials that participants dichotomized Twiggums and Zuttles for analyses (see *Table C.1* for individual trial level data). Higher scores reflect a greater frequency of inferring that the uncharacterized group had opposite traits to the characterized group. We removed one 8-year-old from analyses because she both dichotomized and aligned the groups on every trial (e.g., stating that the uncharacterized group both liked pears and hated pears). Although this is a potentially logical response, we excluded her because we could not represent her decisions within our coding scheme. One 5-year-old also followed this pattern for only one characteristic (i.e., stating that the uncharacterized group was both good and bad at playing drums) so we calculated his proportion score out of seven rather than eight trials (N = 180).

General Procedure. Participants were tested individually in a quiet room as part of a larger study on social reasoning.[21] Participants completed the social categorization task (as reported in Goldfarb et al., 2017) prior to the uncharacterized group task. Children received $10.00 and adults were given course credit for the approximately one-hour session.

Results and Discussion

We conducted analyses in RStudio (R Version 3.6.1; R Core Team, 2017). We tested the effects of generic language and age on the proportion of trials that participants dichotomized the two groups (*Table 7*, *Figure 8* and *9*). Language and age were between-subjects factors. A 2

[21]Participants also responded to questions aimed at addressing their essentialist beliefs, working memory, and fairness and procedural justice judgments. These measures will be analyzed in separate manuscripts.

(language: specific, generic) x 3 (age: 5-, 8-year-olds, adults) univariate analysis of variance (ANOVA) resulted in main effects for language, $F(1, 174) = 7.65$, $p = .006$, $\eta_p^2 = .04$, and age, $F(2, 174) = 8.90$, $p < .001$, $\eta_p^2 = .09$. The Language x Age interaction was not significant, $F(2, 174) = 1.34$, $p = .265$, $\eta_p^2 = .02$.

As anticipated, participants in the generic-language condition dichotomized Zuttles and Twiggums more frequently than those in the specific-language condition ($p = .006$). Still, participants in *both* language conditions judged the two groups to have opposite preferences and skills more often than would be expected by chance (.50; generic: $t[82] = 6.66$, $p < .001$, $d = 0.73$; specific: $t[96] = 3.70$, $p < .001$, $d = 0.38$). Consistent with our predictions, Tukey's HSD comparisons showed that adults dichotomized the groups more frequently than did 5-year-olds ($p < .001$). Eight-year-olds did not differ from either 5-year-olds ($p = .065$) or adults ($p = .111$). Whereas adults ($t[52] = 7.18$, $p < .001$, $d = 0.99$) and 8-year-olds ($t[60] = 4.43$, $p < .001$, $d = 0.57$) dichotomized the two groups more frequently than would be expected by chance, 5-year-olds responded at chance ($t[65] = 1.69$, $p = .096$, $d = 0.21$).

Table 7. Chapter 3: *Study 1 Means (SDs) of Proportion of Trials Dichotomized by Age Group and Language Condition*

	5-year-olds	8-year-olds	Adults	Across Age
Specific Language	.55 (0.29)	.60 (0.34)	.73 (0.36)	.63 (0.33)
Generic Language	.58 (0.36)	.79 (0.33)	.90 (0.15)	.74 (0.33)
Across Language	.57 (0.32)	.70 (0.35)	.80 (0.30)	.68 (0.34)

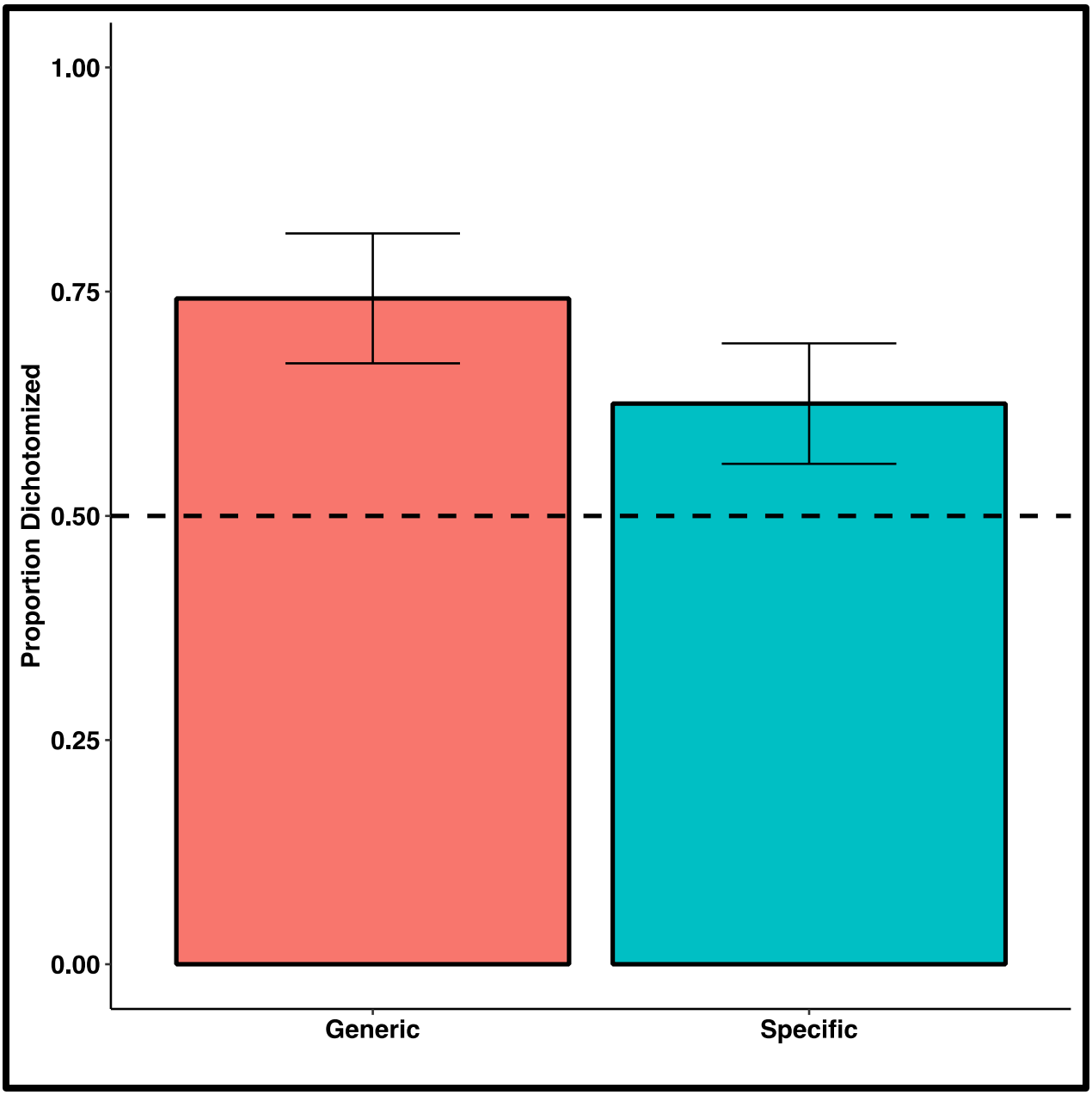

Figure 8. Chapter 3: Study 1 Proportion of trials that participants dichotomized the Zuttles (characterized group) and Twiggums (uncharacterized group) by language condition.
Note. Error bars are 95% confidence intervals. Dashed line signifies chance.

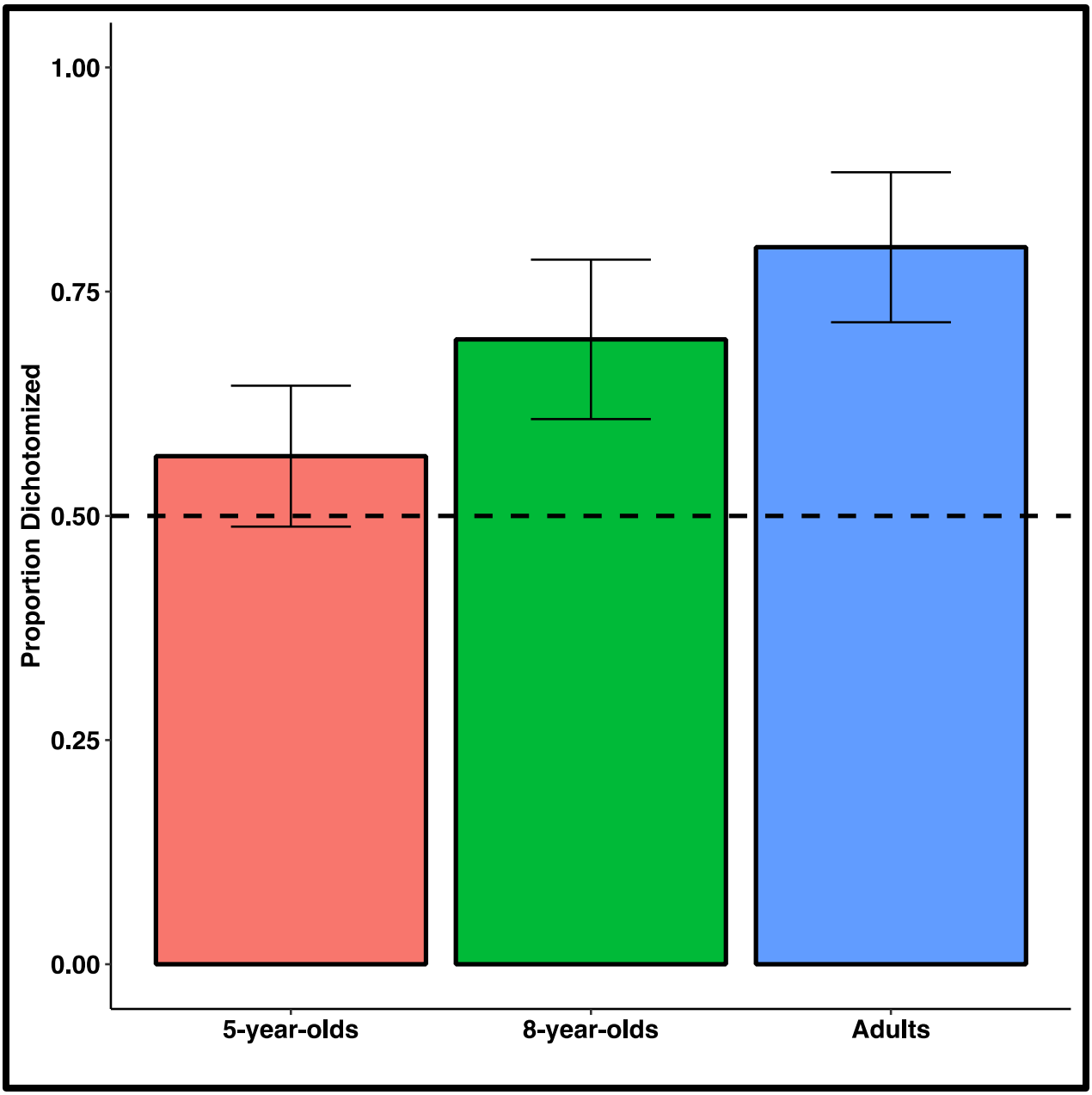

Figure 9. Chapter 3: Study 1 Proportion of trials that participants dichotomized the Zuttles (characterized group) and Twiggums (uncharacterized group) by age group.
Note. Error bars are 95% confidence intervals. Dashed line signifies chance.

Our results indicate that by eight years of age, individuals exhibit a dichotomizing heuristic when forming impressions of new social groups. Older children and adults used what they knew about one group to infer that the opposite was true of the other group. Although participants made this inference in both language conditions, learning about the characterized

group as a whole (generically versus as specific individuals) boosted the frequency that participants expected dichotomies between social groups. These findings are consistent with Moty and Rhodes (2021) in that generic language magnified children's and adults' expectation of social dichotomies, and older children and adults assume dichotomies more than younger children. Still, it is notable that although the dichotomizing heuristic was strong at the level of the individual (e.g., 81% of 8-year-olds in the generic language condition dichotomized the groups on more than 50% of trials), a minority of participants consistently assumed that the two groups would have identical preferences and abilities (see *Table C.1*). This provides assurance that aligning the groups is also possible within our paradigm.

Study 2

Study 1 confirmed the presence of the dichotomizing heuristic, but questions remain as to the circumstances under which it emerges. Prior research has demonstrated that group divisions must serve a functional purpose before people will use these categories in meaningful ways (Bigler, 1995; Bigler et al., 1997). Perhaps our social categorization task, which preceded the uncharacterized group task, acted as a catalyst for the heuristic because it encouraged participants to imagine that Twiggums and Zuttles were "two different kinds"—the explicit starting premise in Moty and Rhodes (2021). In a pilot study for Study 2 (see Online Supplementary Information), we found preliminary evidence that removing the categorization task reduced the dichotomizing heuristic. As a more rigorous test in Study 2, we manipulated (between subjects) whether participants did or did not complete the social categorization task prior to making inferences about the uncharacterized group. This enabled us to test whether repetition separating individuals into two social groups by making characteristic-to-group inferences (i.e., deciding whether each individual is a Zuttle or a Twiggum based on varying

evidence about their preferences and abilities) activates the dichotomizing heuristic.

In Study 1, all participants learned that Twiggums criminalized Zuttles living in the Twiggum Valley (if found, Zuttles would be jailed for 100 days and deported). This evidence of between-group conflict may have boosted the use of the dichotomizing heuristic. Enforced physical and legal boundaries could have signaled to participants that there were perceived incompatibilities between Zuttles and Twiggums. Indeed, intergroup conflict can magnify negative views of outgroup members as well as beliefs about differences (Allport, 1954; Dovidio et al., 2005; Esses et al., 2005). In Study 2, we tested whether intergroup conflict affected adults' tendency to use the dichotomizing heuristic. We manipulated (between subjects) whether the two groups held a negative or positive stance towards each other.

Study 1 focused on categorizing human-like, imaginary creatures. Although this fictional context is a common method in novel group paradigms (see Killen, 2007; Roberts et al., 2017), we sought to test whether participants would apply the dichotomizing heuristic to human social groups. Thus, Study 2 involved novel human groups. As in Study 1, we explicitly stated that the two groups shared the same biological and social origin.

Finally, we examined the scope of the dichotomizing heuristic. Participants in Study 1 may have interpreted the uncharacterized group task as assessing what they "learned" in the social categorization task. If so, then their willingness to expect dichotomies should be confined only to qualities they were featured during the social categorization task. Perhaps, however, the dichotomizing heuristic has even greater reach. In Study 2, we measured whether adults would apply this inference to novel preferences and skills (e.g., if told that Zuttles like "modies" do they infer that Twiggums hate "modies?"). We also examined whether they would even use this heuristic when reasoning about evaluative traits (e.g., if told that Zuttles are "smart" do they infer

that Twiggums are "not smart?"). Indeed, the potential real-world negative consequences of the dichotomizing heuristic would be especially severe if people use it when inferring evaluative traits that signal morals or virtues of an uncharacterized group.

Given these multiple changes needed to systematically unpack the conditions that give rise to the dichotomizing heuristic as well as its potential scope and boundaries, we shifted to an exclusive focus on adults for Studies 2 and 3. We return to the necessity of understanding the developmental trajectory of the dichotomizing heuristic in the General Discussion.

Method

Participants

Participants were 214 undergraduate students (M = 21.13 years, SD = 2.43 years, 105, males, 109 females). Included participants were fluent in comprehending and speaking English and were typically developing (via self-report). The sample was 55% Asian, 2% Black or African American, 17% Hispanic/Latino, 14% White, 8% Multiracial or Multiethnic, and 4% other races or ethnicities. We set the sample size for each cell at 53 participants per cell (N = 212) based on Simonsohn's (2015) recommendation of multiplying the original sample (n = 21 adults in the generic language condition of Study 1) by 2.5. This sample size (53 participants per condition) gave us greater than 99% power to detect a d = 2.76 (the effect size of the dichotomizing heuristic in the adults in the generic language condition in Study 1) for a one-sample t-test (the key test of whether participants in each condition exhibit the dichotomizing heuristic); for the same analysis, it also gives us 80% power to detect a d > 0.39. We stopped data collection the day this goal was met. We included three attention checks that appeared randomly during the dichotomizing phases of the study ([1] Please select the number 4; options: 4, 8; [2] Please select the number 3; options: 1, 3; [3] Please select the number 7; options: 7, 4).

We excluded four participants who were not 100% accurate. The final sample included 210 participants. Participants received course credit. Data were collected from June 2019 through October 2019. This study was approved by the Internal Review Board at the [BLINDED]: Protocol Number: 1031991. The sample size (including exclusions), method, and planned analyses were pre-registered on Open Science Framework (https://osf.io/nw5jx/?view_only=811245f5693a4b3f8f3112b5a7dec7d3).

Measures and Procedure

This study followed a 2 (categorization: categorization, no-categorization) x 2 (stance: negative, positive) x 3 (type: initial, novel, evaluative) design. Categorization was a between-subjects factor that differed in whether participants did (categorization) or did not (no-categorization) complete a social categorization task before the dichotomizing measure. Stance was a between-subjects factor that differed in whether there was conflict (negative) or goodwill (positive) between the two groups. Type was a within-subjects factor where participants responded to three phases of the dichotomizing measure (see details below).

Narrative and Social Categorization Task. All participants were tested individually in a quiet room at a university laboratory. After completing informed consent and basic demographic information, participants responded to a series of tasks via Qualtrics Survey Software (Qualtrics, Provo, UT). Participants learned about a group of humans called Twiggums who lived on an island. They were also told that some Twiggums left Twiggum Island to form a new group called Zuttles who lived on Zuttle Island. Participants then learned about 12 Zuttle characteristics in generic language (i.e., their preferences and abilities; e.g., Zuttles like pears; Zuttles are bad at playing soccer). These were the same characteristics used in Study 1. All participants were given a piece of paper (to put next to the computer) that listed all of the learned

Zuttle characteristics (we used four different blocked orders for this memory card).

Participants in the two categorization conditions then completed the categorization task. As with Study 1, they were told three qualities about each of a series of individuals, and they had to judge whether each person was a Zuttle or a Twiggum. Participants in the Categorization/Negative (Cat/Neg) condition were told that there was a law prohibiting Zuttles from living on Twiggum Island: They sorted individuals to determine who would be deported back to Zuttle Island. In contrast, participants in the Categorization/Positive (Cat/Pos) condition learned that there was a law allowing Zuttles to live on Twiggum Island: They sorted individuals to determine who would be invited to a "welcome" party. This social categorization task, including randomization by evidence type, was identical to Study 1 except that the information was displayed in a written format rather than using a narrated story with pictures. Participants in the no-categorization conditions (No-Categorization/Negative [NoCat/Neg]; No-Categorization/Positive [NoCat/Pos]) read the same information as those in the categorization conditions (i.e., the 12 Zuttle characteristics, the law either prohibiting or welcoming Zuttles to live on Twiggum Island), but they did not complete the social categorization (sorting) task.

Uncharacterized Group Task. All participants then completed three questioning sets where they inferred the traits of the Twiggums. Questions followed a similar format to that of Study 1: Participants were asked to judge whether Twiggums had various preferences, abilities, and traits. These questioning sets followed the same fixed order for all participants.

First, the *initial set* included the 12 characteristics that were featured in the Zuttle familiarization phase and appeared in the categorization task. Thus, this set is a close replication of the uncharacterized group judgment task used in Study 1 (e.g., "Do you think that Twiggums like pears or hate pears?" Answer choices: Twiggums like pears; Twiggums hate pears). Within

this questioning set, responses showed excellent internal consistency across conditions (Cronbach's alpha = .98, 95% CI [.97, .98]). For analyses, we created a proportion of initial trials dichotomized (see *Table C.2* for individual trial level data).

Second, the *novel set* included 12 preferences and abilities described using novel words presented for the first time (e.g., liking "blickets"). The use of novel words ensured that participants could not rely on what they had learned during the initial Zuttle familiarization or had applied in the categorization task. Participants responded to four blocks of three questions (presented in random order; e.g., "Here are some foods that you have never heard about. Zuttles like blickets, modies, and kifs. So, what do you think about Twiggums? Do you think Twiggums like blickets or hate blickets?" Answer choices: Twiggums like blickets, Twiggums hate blickets). Within this questioning set, responses showed excellent internal consistency across conditions (Cronbach's alpha = .97, 95% CI [.96, .98]). For analyses, we created a proportion of novel trials dichotomized (see *Table C.3* for individual trial level data).

Finally, in the *evaluative set* participants were told 12 evaluative traits about Zuttles and made inferences about Twiggums. Participants learned about 4 positive traits (smart, creative, honest, funny), 4 equivocal traits (shy, confident, superstitious, liberal), and 4 negative traits (mean, moody, messy, lazy). For each trait, participants were randomly assigned to learn that Zuttles either did (two of each valence: positive, equivocal, negative) or did not (two of each valence) possess the trait (e.g., "Zuttles are smart. So, what do you think about Twiggums?" Answers: Twiggums are smart; Twiggums are not smart). Traits were presented in randomized order. Within this questioning set, responses showed excellent internal consistency across conditions (Cronbach's alpha = .93, 95% CI [.92, .94]).[22] For analyses, we created a proportion

[22]After responding to the key measures reported in this manuscript (i.e., the social categorization task and the uncharacterized group judgment task), participants answered some exploratory pilot questions (i.e., how similar

of evaluative trials dichotomized (see *Table C.4* for individual trial level data).

Results and Discussion

We examined whether the proportion of trials that participants dichotomized the characterized and uncharacterized groups differed by categorization condition (categorization vs. no categorization task), stance (negative vs. positive intergroup relations), and judgment type (initial vs. novel vs. evaluative traits). Analyses were conducted in RStudio (version 3.6.1; R Core Team, 2019). See Appendix C for relations between the social categorization task and the uncharacterized group task.[23]

We ran a 2 (categorization: categorization, no categorization) x 2 (stance: negative, positive) x 3 (type: initial, novel, evaluative) mixed-effects ANOVA on the proportion of trials that participants dichotomized the groups. Categorization and stance were between-subjects factors and type was a within-subjects factor. This analysis resulted in a main effect for categorization, $F(1, 206) = 195.48$, $p < .001$, $\eta_p^2 = .49$, and type, $F(2, 412) = 6.09$, $p = .002$, $\eta_p^2 = .03$, qualified by Categorization x Type, $F(2, 412) = 17.51$, $p < .001$, $\eta_p^2 = .08$, and Categorization x Stance x Type interactions, $F(2, 412) = 3.50$, $p = .031$, $\eta_p^2 = .02$.

As shown in *Table 8* and *Figure 10*, participants in the categorization conditions dichotomized the groups more often than those in the no-categorization conditions for each of the judgment types (initial, novel, evaluative) and regardless of intergroup stance (*p*s < .001). Although stance had no effect in the categorization conditions (Cat/Neg = Cat/Pos for initial,

participants viewed themselves to Zuttles; how similar participants viewed themselves to Twiggums; how similar participants viewed Zuttles and Twiggums; and what they thought some of the novel words meant).

[23]Although it was not a primary goal to replicate the findings from Goldfarb et al. (2017), the design of the current study afforded this opportunity. Conceptually replicating Goldfarb et al. (2017), participants categorized individuals at comparable rates regardless of the consequence (i.e., deportation vs. invitation to a "welcome" party). Also replicating Goldfarb et al. (2017), we found that participants were more likely to categorize individuals as Zuttles the more evidence matched what they learned about Zuttles, and adults were willing to categorize an unknown individual as a Zuttle at above-chance levels as long as at least one piece of evidence matched what they had learned about Zuttles. See Appendix C for analytic details.

novel, and evaluative traits, *p*s > .647), negative intergroup relations boosted dichotomizing responses in the no-categorization conditions for the initial (NoCat/Neg > NoCat/Pos, $p = .009$) and novel questioning sets (NoCat/Neg > NoCat/Pos, $p = .008$). Participants in the categorization conditions (Cat/Neg, Cat/Pos) dichotomized the two groups more frequently than would be expected by chance for all judgment types (see chance analyses below), but they did so more frequently for initial > novel > evaluative (*p*s < .045; but for Cat/Pos: initial = novel, $p = .064$). In contrast, participants in the no-categorization conditions dichotomized the groups at comparably low rates and below chance (see below) regardless of judgment type (*p*s > .230; but for NoCat/Pos: evaluative > initial, novel, *p*s < .018).

Participants who completed the social categorization measure prior to the uncharacterized group task (Cat/Neg, Cat/Pos) assumed that the two groups had opposing traits (compared to chance [.50]) across all judgment types (initial, novel, evaluative; Cat/Neg: *t*[51]s > 3.18, *p*s < .003, *d*s > 0.44; Cat/Pos: *t*[51]s > 2.90, *p*s < .005, *d*s > 0.40). In contrast, participants in the no-categorization conditions (NoCat/Neg, NoCat/Pos) dichotomized the groups below chance across all judgment types (NoCat/Neg: *t*[53]s > 3.74, $p < .001$, *d*s > 0.51; NoCat/Pos: *t*[51]s > 7.25, *p*s < .001, *d*s > 1.01). These effects also consistently occurred at the individual level (see *Tables C.2, C.3, C.4*). For example, whereas 85% of participants in the Cat/Neg and Cat/Pos conditions dichotomized the groups on more than 50% of trials, only 24% of participants in the NoCat/Neg and 10% of participants in NoCat/Pos conditions followed this pattern.

Table 8. Chapter 3: Study 2 Means (SDs) of Proportion of Trials Dichotomized

	Categorization-Negative (n = 52)	Categorization-Positive (n = 52)	No-Categorization-Negative (n = 54)	No-Categorization-Positive (n = 52)	All Conditions
Initial	.83 (0.29)	.85 (0.31)	.28 (0.39)	.11 (0.22)	0.51 (0.45)
Novel	.74 (0.36)	.77 (0.34)	.30 (0.38)	.14 (0.23)	0.49 (0.43)
Evaluative	.66 (0.35)	.64 (0.35)	.26 (0.30)	.23 (0.26)	0.44 (0.37)
All	.74 (0.29)	.75 (0.28)	.28 (0.32)	.16 (0.18)	0.48 (0.38)

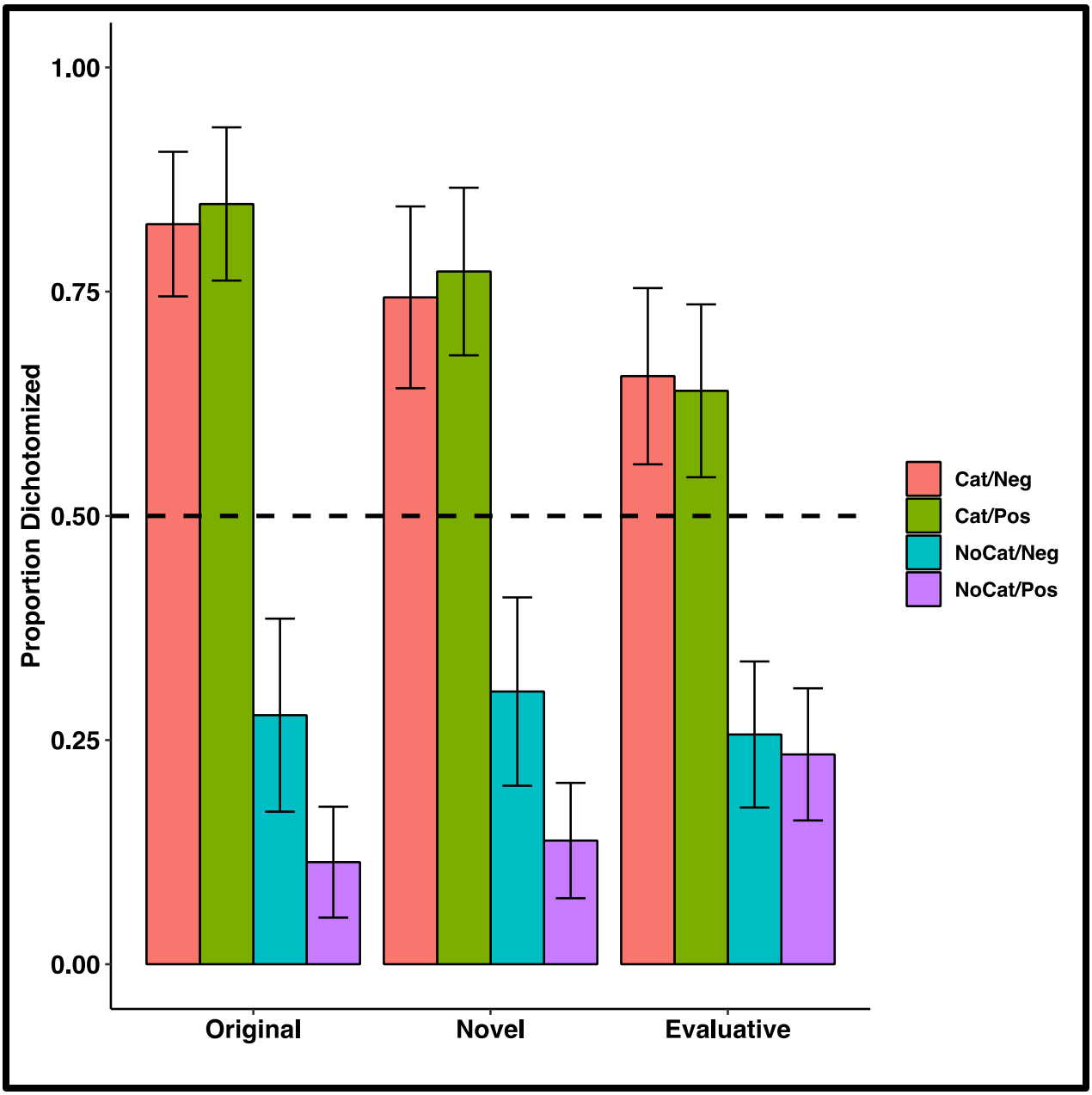

Figure 10. Chapter 3: Study 2 Proportion of trials that participants dichotomized the Zuttles (characterized group) and Twiggums (uncharacterized group) by type (Initial, Novel, Evaluative), categorization (categorization, no-categorization), and stance conditions (positive, negative). Note. Error bars are 95% confidence intervals. Dashed line signifies chance. Cat/Neg = Categorization/Negative; Cat/Pos = Categorization/Positive; NoCat/Neg = No-Categorization/Negative; NoCat/Pos = No-Categorization/Positive.

Study 2 clarified the factors that lead people to use a dichotomizing heuristic when forming impressions of social groups: Adults only displayed this rule-of-thumb shortcut at high

frequency when they perceived the two groups as separate via making characteristic-to category inferences during the social categorization task. In contrast, animosity between the groups was not necessary. Beyond illustrating what catalyzes this cognitive shortcut, we also documented its expansive range. That is, we discovered that adults not only assumed that the uncharacterized group (Twiggums) differed from the characterized group (Zuttles) when reasoning about the initial set of preferences and abilities that they applied during the social categorization task (e.g., whether a group likes pears), but also when considering novel preferences and abilities (e.g., whether a group likes modies) as well as evaluative traits (e.g., whether a group is lazy).

Study 3

Study 2 demonstrated that repetitively attempting to sort individuals into separate social groups based on evidence about each individual's characteristics was necessary for adults to show the dichotomizing heuristic when later asked to make inferences about the preferences and skills of an uncharacterized group. It is possible, however, that the categorization task would have carried less weight if participants had not been told that Zuttles and Twiggums shared an identical biological and social origin as well as looked the same. Thus, in Study 3, we tested whether making the social groups functionally separate from the start would be sufficient to elicit the dichotomizing heuristic, even in the absence of the categorization task.

Study 3 also addressed two additional methodological limitations. In Studies 1 and 2 as well as in Moty and Rhodes (2021), the narratives highlighted binary contrasts: There were two groups (i.e., Zuttles vs. Twiggums) and participants made binary choices (e.g., They judged whether Twiggums were either good or bad at basketball). Perhaps these binary cues augmented the dichotomizing heuristic. Designing the prior studies in these ways was sensible because we were interested in capturing what participants assumed about two groups when forced to choose.

From a methodological and theoretical perspective, however, it is important to examine how these factors influence the dichotomizing heuristic. In Study 3, therefore, we allowed participants to make scaled (as opposed to binary) judgments, and we manipulated between subjects whether participants learned that there were many groups versus only two groups who lived in the region.

Because we found that participants in Study 2 were willing to dichotomize groups even when reasoning about evaluative traits (e.g., whether a social group was smart or lazy), we tested a new potential boundary condition. That is, in Study 3 we measured adults' use of the dichotomizing heuristic when the characterized group (Zuttles) had preferences that one would expect most humans to have (e.g., Zuttles dislike getting sick; Zuttles like feeling happy). We will refer to these characteristics as *universal*. Here, to infer that the uncharacterized group (Twiggums) had opposite characteristics would require participants to judge that this group held extremely atypical preferences (e.g., they like getting sick; they dislike feeling happy).

We also investigated a possible outcome related to the dichotomizing heuristic. In particular, we were curious about whether the tendency to dichotomize is related to beliefs about broader intergroup relations. To do this, we explored whether individuals who assumed greater differences between Zuttles and Twiggums also expected that the two groups lived more segregated (versus integrated) lives in their social communities. We included questions about whether Zuttles and Twiggums lived in the same neighborhoods, went to the same schools, were friends, were romantically involved with each other, and went to the same doctors. Assessing beliefs about the Twiggum/Zuttle society is also of interest because it speaks to the contexts under which people might utilize the dichotomizing heuristic. For example, do adults only apply the dichotomizing heuristic when they assume that the groups live separate lives or do they still make this cognitive shortcut even when they think that a society is well integrated?

Method

Participants

Participants were 212 undergraduate students (M = 20.20 years, SD = 2.25 years, 43 males, 162 females, 1 gender non-binary).[24] Included participants were fluent in comprehending and speaking English and were typically developing (via self-report). The sample was 39% East Asian (e.g., Chinese, Korean, Japanese), 1% Native Hawaiian or Other Pacific Islander, 11% South Asian (e.g., Indian, Pakistani), 12% South East Asian (e.g., Thai, Vietnamese, Filipino), 8% White, and 10% another race or ethnicity or multiracial.[25] We set the sample size at 53 participants per cell (N = 212) to remain consistent with Study 2. In addition, 53 participants per condition gives us 82% power to detect a d = 0.40 or greater (the smallest effect detected in Study 2) for a one-sample t-test. We stopped data collection the day this goal was met. We included five attention checks throughout the study (e.g., Please select the number 3; options: 1, 2, 3, 4, 5). We excluded 6 participants who missed two or more of these checks. The final sample included 206 participants. Participants received course credit.

Data were collected in February 2021. This study was approved by the Internal Review Board at the [BLINDED]: Protocol Number: 1031991. The sample size (including exclusions), method, and planned analyses were pre-registered on the Open Science Framework (https://osf.io/73gdh/?view_only=b81c4524835c4aed99e0ff9192d9a591).

Measures and Procedure

This study followed a 2 (size: two, many) x 2 (categorization: categorization, no-categorization) x 4 (type: initial, novel, evaluative, universal) design. Size was a between-

[24]Three additional participants consented to the study but did not complete it (i.e., they were never assigned to a condition). One of these participants provided some basic demographic information, but did not respond to any of the other variables. The other two participants, consented but then responded to no additional variables.
[25]Between Study 2 and Study 3 we updated our basic demographics to be more specific.

subjects factors that varied in whether participants learned that two groups lived on the island (Two) or whether many groups lived on the island (Many). Categorization was a between-subjects factor that differed in whether participants did (Cat) or did not (NoCat) complete a social categorization task before the dichotomizing measure. Type was a within-subject factor where participants engaged in four phases of the dichotomizing task which each inquired about different types of characteristics (initial, novel, evaluative, and universal).

Narrative and Social Categorization Task. All participants were tested online via Qualtrics Survey Software (Qualtrics, Provo, UT) due to the restrictions on in-person testing caused by the Covid-19 pandemic. After completing informed consent and basic demographic information, participants responded to the main tasks. All participants learned about an island where groups of humans lived. Half of the participants learned that two groups of humans lived on the island (Two-Group Condition: "Imagine that there are two groups of humans who live on an island. One of the groups is called Zuttles. The other group is called Twiggums"). In the other condition, participants learned that many groups of humans lived on the island (Many-Group Condition: "Imagine that there are many groups of humans who live on an island. One of the groups is called Zuttles."). We included a comprehension check ("So, how many groups of humans live on the island?" Choices: 1, 2, many). Participants were not able to move on until they answered this question correctly. In both conditions, we avoided using any "difference" language so as to not cue that the groups differed beyond having separate category labels.

Next, all participants learned about the 12 Zuttle characteristics (i.e., their preferences and abilities as in Studies 1 and 2; e.g., Zuttles like pears). Half of the participants then completed the categorization task. In the two-group version (Two-Group/Categorization [Two/Cat]), the categorization task was the same as Study 2. For participants in the many-groups

version (Many-Group/Categorization [Many/Cat]), the only difference was that participants were asked to determine whether each individual was a Zuttle or "someone from another group." Participants in the no-categorization condition did not complete the categorization task (Two-Group/No-Categorization [Two/NoCat]; Many-Group/No-Categorization [Many/NoCat]).

Uncharacterized Group Task. All participants then completed four questioning sets where they made inferences about the preferences, skills, and traits of the uncharacterized group, the Twiggums. Participants in the Two Group Condition (Two/Cat, Two/NoCat) were introduced to the task in the same way as participants in Study 2. Participants in the Many Group Condition (Many/Cat, Many/NoCat) were told that they would be answering questions about one of the other groups, Twiggums. These were similar questions to Study 2 except that participants made judgments about Twiggums' preferences, abilities, and traits using a 5-point scale rather than making a binary choice. These questioning sets followed the same fixed order for all participants.

First, participants responded to the *initial set* of the 12 characteristics that were included in the Zuttle familiarization narrative (all conditions) and also appeared in the categorization task (Two/Cat, Many/Cat). On each trial, participants were reminded of the Zuttles' characteristic, and then they made a judgment about Twiggums (e.g., "Remember…Zuttles like pears. So, what do you think about Twiggums?", 5-point scale: ranging from Twiggums DISLIKE pears to Twiggums LIKE pears, midpoint = Twiggums are NEUTRAL in this preference).

Second, participants responded to the *novel set* of 12 preferences and abilities described using novel words (e.g., being good at playing daxes). Participants responded to four blocks of three questions (presented in random order; e.g., "Here are some activities that you have never heard about. Zuttles are good at playing lups, daxes, and nafs. So, what do you think about Twiggums?", 5-point scale: Twiggums are BAD at playing lups to Twiggums are GOOD at

playing lups, midpoint = Twiggums are NEUTRAL in this skill).

Third, in the *evaluative set* participants were told 12 evaluative traits about Zuttles and made inferences about Twiggums. Participants learned about 4 positive traits (smart, creative, honest, funny), 4 equivocal traits (shy, confident, superstitious, liberal), and 4 negative traits (mean, moody, messy, lazy). For each trait, participants were randomly assigned to learn that Zuttles either did (two of each valence: positive, equivocal, negative) or did not (two of each valence) possess the trait (e.g., "Zuttles are smart. So, what do you think about Twiggums?", 5-point scale: Twiggums ARE NOT smart to Twiggums ARE smart, midpoint = Twiggums are NEUTRAL in this trait). Traits were presented in randomized order.

Fourth, in the *universal set* participants responded to 12 preference trials that involved Zuttles holding universal attitudes (e.g., Zuttles dislike getting sick). Participants learned that Zuttles had 6 universal dislikes (e.g., Zuttles dislike the sound of nails on a chalkboard) and 6 universal likes (e.g., Zuttles like having fun). Participants then made judgments about Twiggums on a 5-point scale (e.g., "So, what do you think about Twiggums?", 5-point scale: Twiggums DISLIKE getting sick to Twiggums LIKE getting sick, midpoint = Twiggums are NEUTRAL in this preference). These items were presented in a random order.

We used a 5-point scale for several reasons. It mirrored prior studies in that participants had to make a judgment about Twiggums' characteristics. It also allowed us keep the same training (i.e., the items that Zuttles liked/disliked and were good/bad at),[26] limiting the number of changes we made from prior studies (maximizing what we could learn in this study). Participants could pick the scale midpoint instead of being forced to pick one side of the scale creating a more stringent test of the dichotomizing heuristic. We labeled the midpoint "neutral in this

[26]The only exception to this was that in the current study participants learned that Zuttles either liked or disliked certain foods rather than learning that they either liked or hated certain foods.

preference/skill/trait" to limit ambiguity about the scale values. Moreover, dichotomizing the groups on this scale required participants to switch sides of the scale depending on the trial. That is, for half of the trials, dichotomizing required participants to pick a "1" or a "2" on the scale whereas for the other half of trials, dichotomizing would lead participants to pick a "4" or a "5."

Coding and scoring. For all questioning sets, the items were rescored so that higher scores indicated a greater difference between the characteristics of Zuttles and the characteristics of Twiggums. For all questioning sets, responses showed excellent internal consistency (initial: Cronbach's alpha = .91, 95% CI [.89, .93]; novel: Cronbach's alpha = .95, 95% CI [.93, .96]; evaluative: Cronbach's alpha = .91, 95% CI [.89, .93]; universal: Cronbach's alpha = .94, 95% CI [.93, .95]). We also rescored the judgments to approximate a dichotomous choice. That is, participants received a score of 1 for every trial that they chose a 4 or 5 (i.e., chose the opposite side of the scale for Twiggums from what they were told about Zuttles) and a score of 0 for every trial that they chose a 1, 2, or 3 on the scale (i.e., chose the same side of the scale for Twiggums as they were told about Zuttles or chose the neutral). See *Tables C.5, C.6, C.7,* and *C.8,* individual level trial data. Including the neutral scale point as a *failure* to dichotomize provides a more conservative test of the dichotomizing heuristic in that we required participants to actually endorse that the uncharacterized group (Twiggums) held *opposite* characteristics.

Proximity Task. At the end of the study, participants responded to questions examining their perceptions of the island where Zuttles and Twiggums live. We were interested in whether participants believed that Zuttles and Twiggums lived segregated or integrated lives on the island. Participants responded to a series of questions that required a binary response: (1) When you picture the island where Zuttles and Twiggums live, Zuttles likely… (a) Only live in neighborhoods with Zuttles or (b) Live in neighborhoods with both Zuttles and Twiggums; (2)

When you picture the island where Zuttles and Twiggums live, Zuttle children likely… (a) Only go to school with Zuttles or (b) Go to school with both Zuttles and Twiggums; (3) When you picture the island where Zuttles and Twiggums live, Zuttles are likely… (a) Only friends with Zuttles or (b) Friends with both Zuttles and Twiggums; (4) When you picture the island where Zuttles and Twiggums live, Zuttles are likely… (a) Only romantically involved with Zuttles or (b) Romantically involved with Zuttles or Twiggums; (5) When you picture the island where Zuttles live, Zuttles likely… (a) Only go to doctors and hospitals used by Zuttles or (b) Go to doctors and hospitals used by both Zuttles and Twiggums. Participants viewed questions in a random order (response options were also randomized). These items showed good internal consistency (Cronbach's alpha = .85, 95% CI [.82, .88]). For analyses, we created a score of the proportion of trials that participants judged that the two groups lived segregated lives.

Results and Discussion

Results are divided into two sections. First, we assessed participants' responses to the uncharacterized group task, both their scaled ratings and their rescored responses that approximated a dichotomous choice. Second, we explored the data from the proximity task. Analyses were conducted in RStudio (version 3.6.1; R Core Team, 2019). See Appendix C for analyses on the social categorization task alone.

Uncharacterized Group Task

We ran a 2 (size: two groups, many groups) x 2 (categorization: categorization, no categorization) x 4 (type: initial, novel, evaluative, universal) mixed-effects ANOVA on the 5-point ratings (the higher the score the greater the difference between Twiggums and Zuttles). Size and categorization were between-subjects' factors and type was a within-subjects' factor. This analysis resulted in a main effect for categorization, $F(1, 202) = 14.41$, $p < .001$, $\eta_p^2 = .07$,

and type, $F(3, 606) = 121.25$, $p < .001$, $\eta_p^2 = .38$ (see *Table 9*). There were no other significant effects (including any for group size), Fs < 3.42, ps > .066, η_p^2s < .02. Thus, conceptually replicating and extending Study 2, participants in the categorization conditions assumed greater differences between the groups than did participants in the no categorization conditions ($p <$ .001). In addition, participants expected greater differences between Twiggums and Zuttles when making inferences about initial, novel, and evaluative characteristics as compared to universal characteristics (ps < .001; all other comparisons: ps > .382).

We used the rescored data (see Coding and Scoring) to examine whether participants dichotomized the groups more frequently than would be expected by chance. We ran a two-tailed t-test compared to chance (.40). Chance is .40 because on any given trial there was a 40% possibility of choosing a 4 (.20) or a 5 (.20) at random (.20 + .20 = .40).[27] Participants in both conditions dichotomized the groups more often than would be expected by chance for initial (e.g., Zuttles like apples), novel (e.g., Zuttles like modies), and evaluative items (e.g., Zuttles are smart; Cat: *M*s > .55, *t*[102]s > 4.26, *p*s < .001, *d*s > 0.42; NoCat: *M*s > .48, *t*[102]s > 2.30, *p*s < .023, *d*s > 0.23; *Figure 11*). Thus, the categorization task was unnecessary to create the heuristic when the two groups were labeled as distinct from the start and had no stated common origin. We also found that participants dichotomized Twiggums' and Zuttles' initial, novel, and evaluative traits above chance irrespective of whether they learned that there were two or many groups on the island (Two Groups: *M*s > .56, *t*[103]s > 3.95, *p*s < .001, *d*s > 0.39; Many Groups: *M*s > .48, *t*[101]s > 2.26, *p*s < .026, *d*s > 0.22). Therefore, we found no evidence that use of the dichotomizing heuristic is dependent on there being an explicit contrast between two groups.

[27]We had planned to conduct these t-tests at the Categorization x Size level, but because this interaction was not significant, it was more appropriate to examine the level of the main effects. *Table 3* shows the means and standard deviations by each condition.

We did, however, discover an important and reasonable boundary condition for the dichotomizing heuristic. Participants in the categorization, no categorization, two groups, and many groups conditions assumed that the groups would be aligned in universal preferences (e.g., they both like feeling happy; *M*s < .32, |*t*|s > 2.07, *p*s < .041, *d*s > 0.20). Said another way, participants were unwilling to view Twiggums as opposite to Zuttles when doing so required that Twiggums had atypical preferences (e.g., Twiggums like getting gum stuck in their hair).[28]

[28]We also conducted one-sample two-tailed t-tests on the proportion of trials that participants selected a 2, 3, 4, or 5 on the scale (i.e., selected a characteristic that differed from Zuttles) compared to chance (.80). For initial, novel, and evaluative characteristics, participants in the categorization, in the two group, and in the many group conditions avoided selecting "1" on the scale (i.e., saying that Twiggums were the same as Zuttles; *M*s > .85, *t*s > 2.18, *p*s < .032, *d*s > 0.22), but participants in the no-categorization condition did not avoid selecting "1" more frequently than would be expected by chance for initial, novel, and evaluative characteristics (*M*s < .82, *t*s < 0.65, *p*s > .515, *d*s < 0.06). In contrast, participants in the both categorization conditions and in the both group conditions chose "1" more frequently than would be expected chance for universal items (*M*s < .66, *t*s > 3.74, *p*s < .001, *d*s > 0.37).

Table 9. Chapter 3: Study 3 Means (SDs) Dichotomizing Ratings, Proportion of Trials Dichotomized, and Proximity Questions

	Two-Categorization (n = 53)	Two-No-Categorization (n = 51)	Many-Categorization (n = 50)	Many-No-Categorization (n = 52)	All Conditions
			Dichotomizing Ratings		
Initial	4.01 (0.86)	3.48 (1.06)	3.69 (0.90)	3.37 (1.04)	3.64 (0.99)
Novel	3.89 (1.03)	3.42 (1.21)	3.52 (0.93)	3.33 (1.12)	3.54 (1.09)
Evaluative	3.96 (0.94)	3.29 (1.11)	3.43 (0.79)	3.40 (0.96)	3.52 (0.99)
Universal	2.95 (1.36)	2.03 (1.05)	2.58 (1.15)	2.23 (0.97)	2.45 (1.19)
All	3.70 (0.82)	3.05 (0.92)	3.30 (0.74)	3.07 (0.80)	3.28 (0.86)
			Proportion of Trials Dichotomized		
Initial	.66 (.33)	.54 (.35)	.56 (.37)	.50 (.33)	.56 (.35)
Novel	.62 (.40)	.49 (.40)	.52 (.39)	.48 (.35)	.53 (.39)
Evaluative	.64 (.36)	.47 (.33)	.46 (.35)	.49 (.33)	.52 (.35)
Universal	.37 (.39)	.18 (.28)	.28 (.34)	.19 (.25)	.26 (.33)
All	.57 (.31)	.42 (.29)	.45 (.30)	.41 (.26)	.47 (.29)
			Proximity Questions		
Neighborhoods	.38 (.49)	.40 (.49)	.37 (.49)	.46 (.50)	.40 (.49)
School	.42 (.50)	.24 (.43)	.27 (.45)	.37 (.49)	.33 (.47)
Friends	.38 (.49)	.34 (.48)	.38 (.49)	.38 (.49)	.37 (.48)
Romantic	.45 (.50)	.32 (.47)	.35 (.48)	.40 (.50)	.38 (.49)
Healthcare	.38 (.49)	.26 (.44)	.35 (.48)	.35 (.48)	.33 (.47)
All	.40 (.42)	.31 (.35)	.35 (.39)	.39 (.37)	.36 (.38)

Note. Dichotomizing ratings range from 1 to 5 (higher scores indicate that participants expect greater differences between Zuttles and Twiggums. Proportion of Trials Dichotomized range from 0 to 1 (higher scores are indicative of dichotomizing; on these trials, chance is .40). Proximity Questions range from 0 to 1 (higher scores are indicative of beliefs that the groups are more segregated).

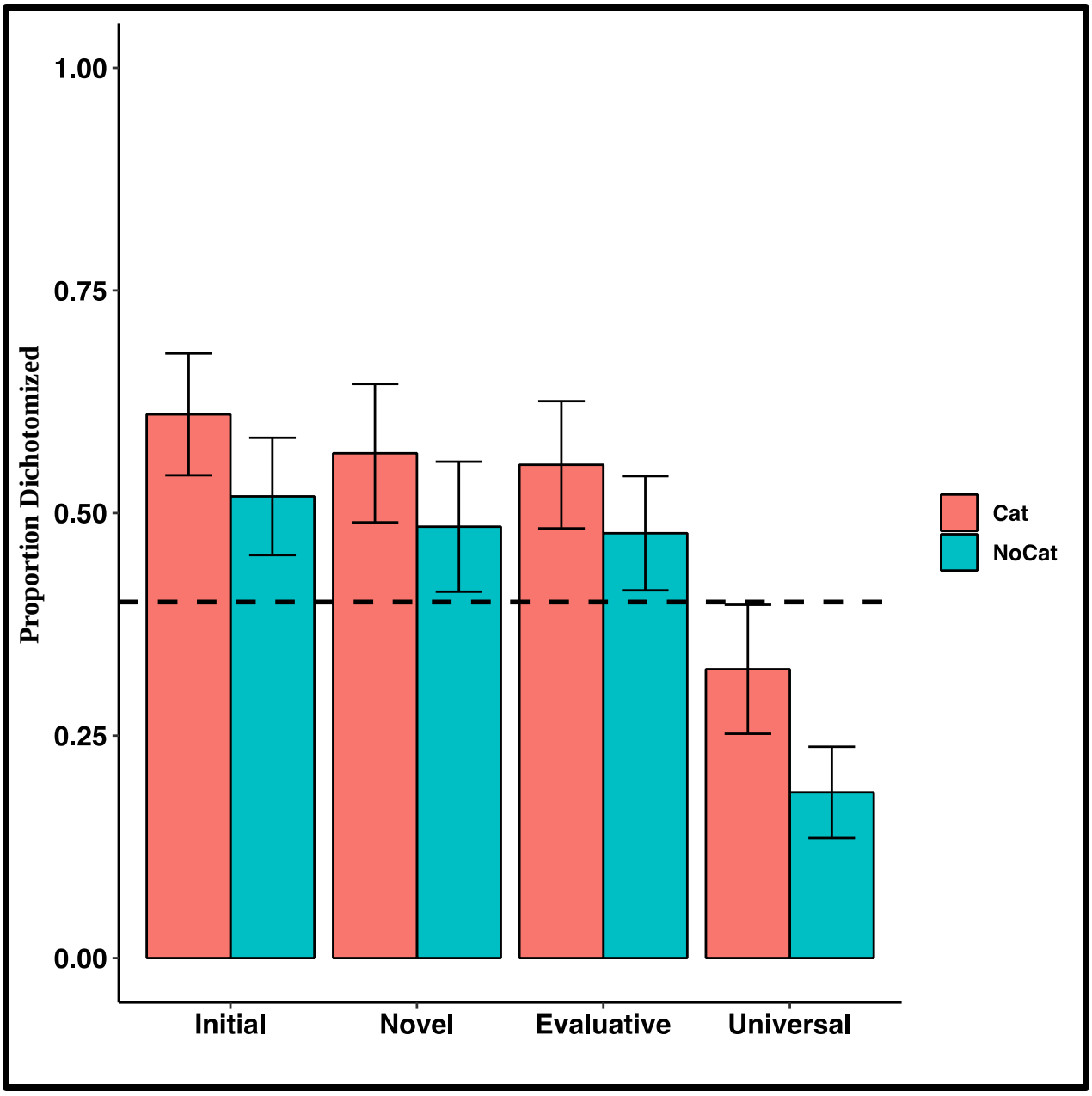

Figure 11. Chapter 3: Study 3 Proportion of trials that participants dichotomized the Zuttles (characterized group) and Twiggums (uncharacterized group) by type (Initial, Novel, Evaluative, Universal) and categorization condition (Categorization, No-Categorization).
Note. Error bars are 95% confidence intervals. Dashed line signifies chance (.40). Cat = Categorization; NoCat = No-Categorization.

Proximity Task

We conducted a 2 (size) x 2 (categorization) ANOVA on the proportion of trials that participants assumed that the groups lived segregated lives. All effects were null, *F*s < 1.62, *p*s >

.205, η_p^2s < .01. Participants expected that Twiggums' and Zuttles' lives were more integrated than segregated (M = .36, 95% CI [.31, .42], t[205] = -5.17, p < .001, d = 0.36).[29] This suggests that participants were willing to use a dichotomizing heuristic even when they viewed the societies as more integrated than segregated. We also examined correlations between dichotomizing ratings (i.e., the 5-point scale) and beliefs about segregation. Controlling for size and categorization conditions, the greater the difference that participants expected between Twiggums and Zuttles (collapsing across judgment type), the more frequently they expected the two groups to live segregated lives (r[202] = .23, 95% CI [.10, .36], p < .001).[30]

General Discussion

Individuals apply a *dichotomizing heuristic* when forming impressions of social groups: They use what they know about one group to infer that the opposite is true about an uncharacterized group. Although participants exhibited this shortcut across the generic and specific language conditions in Study 1, learning about the characterized group as a generic whole boosted its frequency. In Studies 2 and 3 we discovered that conceptualizing the groups as distinct entities (either by engaging in a social categorization task or by initially being told that they are separate groups) is a prerequisite for eliciting the shortcut among adults. Once triggered, however, adults applied the dichotomizing heuristic when making category-to-characteristic inferences about a novel group's benign, novel, and evaluative traits. Studies 2 and 3 also demonstrated that intergroup conflict (i.e., present versus absent) or the number of groups who inhabit the geographical area (i.e., two versus many groups) did not measurably shape use of the

[29]We had planned to run this analysis by condition, but the lack of effects in the ANOVA led us to collapse across conditions (see *Table 3* for means and standard deviations by condition).

[30]This same relation held when we examined each of the individual judgments separately (initial: r[202] = .21,95% CI [.07, .34], p = .003; novel: r[202] = .19, 95% CI [.05, .32], p = .007; evaluative: r[202] = .16, 95% CI [.03, .29], p = .019; universal: r[202] = .17, 95% CI [.04, .30], p = .013).

heuristic. Moreover, adults even exhibited the cognitive shortcut when making scaled rather than binary judgments, and their beliefs did not appear to be fully driven by an assumption that the two groups lived segregated lives. Still, participants exhibited a rational boundary: They did not apply the dichotomizing heuristic when doing so would require them to say that the uncharacterized group held highly unusual preferences (e.g., liking to be sick; disliking to be happy). Below, we expand on these findings and propose suggestions for future research.

Documenting a Dichotomizing Heuristic

When deciding what a new social group is like, a person should ideally take time to make thoughtful evaluations, check for consistencies and inconsistencies in the kinds of characteristics group members express, and avoid making unfounded generalities. Although such practices could help foster fair, accurate impressions, they are time and resource intensive, and as a result, people often use shortcuts. For example, the *availability heuristic*—the ease with which something comes to mind is viewed as a cue to its prevalence (Tversky & Kahneman, 1973)—leads adults to infer that repeatedly observed traits (e.g., laziness), even if only paired with one group member, describe the entire group (Rothbart et al., 1978). Children and adults also expect bidirectional individual-to-group relations: They extend traits from an individual to the group as well as from the group to an individual (Disendruck et al., 2015; Gelman et al., 2010; Rhodes & Gelman, 2008; Sherman et al., 2009).

Here, we discovered another way that individuals streamline social judgments: They use a dichotomizing heuristic. The heuristic operated in the *absence* of direct contact with, prior knowledge of, or learned "facts" about the new group in question. Instead, it only required first conceptualizing the uncharacterized group as separate and distinct from the characterized group. Even without making group differences salient (as in Moty & Rhodes, 2021), the dichotomizing

heuristic emerged across several contexts. Participants exhibited the effect when they learned about the characterized group as a generic whole and as individual group members (Study 1), when there was and was not intergroup conflict (Study 2), when they learned about the characterized group in the context of just one other group or in the context of many other groups (Study 3), and when participants made binary (Study 1 and Study 2) as well as scaled judgments (Study 3). Moreover, adults not only applied it to features that they initially learned about Zuttles that were benign (e.g., liking apples), but they also extended it to novel preferences and skills (e.g., liking modies), and even to evaluative traits that signaled the morals and virtues of a group (e.g., being mean). Notably, among these manipulations, the only context where they did not use the heuristic was when doing so would render Twiggums as having extremely atypical human preferences (e.g., liking spiders crawling on them; disliking winning money).

Several aspects of human cognition likely contribute to the dichotomizing heuristic. Social categorization is powerful: Partitioning individuals into groups can change people's motivations and cognitions (Hugenberg et al., 2010; Young & Hugenberg, 2010; Qian et al., 2019). For example, category labels cue between-group differences in properties, characteristics, and traits (Eiser, 1971; Hamilton & Gifford, 1976; Johnston & Jacobs, 2003; Krueger & Rothbart, 1990; Lawson & Bower, 2014; Master et al., 2012; Tajfel & Wilkes, 1963). Children and adults also utilize speech pragmatics to infer meaning from both what the speaker says and what they leave out (i.e., implied contrasts; Bohn & Frank, 2020; Diesendruck & Markson, 2001; Grice, 1975; Halberda, 2006; Markman et al., 2003; Papfragou & Musolino, 2003). These beliefs about group differences, combined with language pragmatics, likely work in tandem to facilitate the dichotomizing heuristic. In other words, people have general beliefs that distinct categories differ on at least some factors. Given this starting assumption, people may then search for the

features that differentiate groups (Sherman et al., 2009). When a speaker provides information about one group (e.g., Zuttles like apples), the listener may infer that the decision to exclude information about the second group was purposeful (i.e., this must be the key differentiating information). Said another way, the listener may assume that if the information was relevant to both groups then the speaker would have said as much.

Importantly, however, data from Study 3 revealed that the dichotomizing heuristic cannot be reduced to pragmatics alone. In contrast to inferring that Twiggums had opposite characteristics to Zuttles when judging benign, novel, and evaluative traits, participants predicted that Twiggums and Zuttles would *share* preferences considered universal to humans (e.g., they expected both Twiggums and Zuttles to like feeling happy and to dislike getting hurt). This evidence highlights a clear boundary condition for the heuristic—adults are unlikely to apply it when doing so would lead them to make highly improbable assumptions about a social group. Potentially, then, adults' use of the dichotomizing heuristic may be influenced by their beliefs about how common or uncommon a particular characteristic is in a population. Still, many of the characteristics participants consistently dichotomized for initial and evaluative trait sets could also be viewed as widespread (e.g., lots of people like apples). Thus, base rates may only affect the dichotomizing heuristic in extreme boundary cases such as those tested in Study 3. Future studies are needed that manipulate base rates and test dichotomous thinking about social groups.

Catalysts of the Dichotomizing Heuristic

Generic language is pervasive: Its ubiquity crosses the boundaries of language (Chierchia, 1998) and context, including parent-child conversations (Gelman et al., 2008) and news reports (Dukes & Gaither, 2017). Previous studies have shown that hearing about groups in generic versus specific language encourages children and adults to assume greater homogeneity

within a group (Rhodes et al., 2012). Here, we show that generic language also permeates concepts of uncharacterized groups (see also Moty & Rhodes, 2021). That is, generic language (versus specific language) inflated the assumption that two groups must have opposite preferences and abilities despite information only being provided about one of the groups. This type of speech may have made the descriptions of the characterized group appear more exclusive; generic language is deemed more appropriate when a characteristic is distinctive (i.e., characteristic is specific to that group; e.g., lions have manes) versus non-distinctive (i.e., characteristic is widespread across groups; e.g., lions are males; Cimpian et al., 2010; Leslie, 2008). Although dampened, participants who learned about individual Zuttles in specific language condition also dichotomized the two groups. Thus, while generics appears to further cement the use of the shortcut, discussing group members as individuals does not eradicate unsupported inferences about unknown groups (see also Goldfarb et al., 2017).

Across studies, we learned that participants apply the dichotomizing heuristic if they conceptualize the groups as distinct and separate, and there are multiple avenues for forming this impression. When two groups derived from the same biological and social origin and looked identical (Studies 1 and 2), a brief experience of attempting to categorize individuals into two groups based only on information about the preferences and skills of one group led participants to later judge that the two groups must have opposite traits. Another approach is to tell participants that there are two or more social groups (no origin information), describe characteristics about one group, and then ask them to infer what the other group is like (Study 3). This strategy produced the dichotomizing effect even though we avoided language that would cue differences beyond the groups having separate labels (i.e., we didn't use leading words such as "different," "distinct," or "separate" when discussing the groups and provided no evidence

that they had distinguishing physical features). The least subtle pathway to elicit the heuristic is to describe groups are "different kinds" and show that they look different too—enabling children as young as 4.5 years of age to dichotomize groups (Moty & Rhodes, 2021). Indeed, our developmental findings from Study 1 underscore the interpretation that perceiving groups as distinct is a necessary trigger of the dichotomizing heuristic. Consistent with prior work indicating weaker category boundaries and less consistent characteristic-to-category judgments in children 5 years and younger compared to older children and adults (Goldfarb et al., 2017; Lagattuta & Kramer, 2021; Rhodes & Gelman, 2008; Riggs et al., 2014; Shutts, 2015; Sloutsky et al., 2015), the social categorization task may not have sufficiently induced 5-year-olds to treat Zuttles and Twiggums as distinct groups. Young children likely need, as in Moty and Rhodes (2021), to be unambiguously told and shown that members of one group differ from members of another group for them to apply this cognitive shortcut.

Studies 2 and 3 further identified factors that failed to influence adults' use of the dichotomizing heuristic. First, we did not find strong support that intergroup conflict (versus goodwill between groups) mattered: Adults dichotomized groups at equivalent rates regardless of whether Twiggums prohibited versus welcomed Zuttles to their community. Still, there may be contexts in which negative intergroup relations could elicit the shortcut that groups must hold opposite characteristics. For example, in the current research, we used novel groups to which the children and adults did not belong so that neither existing knowledge nor in-group bias (Killen, 2007; Rhodes et al., 2012; Roberts et al., 2017) could shape judgments. Potentially, if participants were members of one of the groups, intergroup animosity could boost dichotomizing rates. Second, Study 3 showed that the number of social groups featured in the learning context (i.e., two versus many) did not affect participants' use of the dichotomizing heuristic. Both

participants in the two-group condition and those in the many-group condition dichotomized the groups, suggesting that the heuristic was not simply produced by an assumption that we (as the researchers) had intentionally set up an explicit contrast between the two groups.

Limitations and Future Directions

The current studies only tested the dichotomizing heuristic using an experimental paradigm featuring novel groups. We think it is likely that people also apply this cognitive shortcut in real-world settings. For example, the dichotomizing heuristic could contribute to why some individuals erroneously assume that the Black Lives Matter movement means that other lives do not matter. Indeed, beliefs about the dichotomous nature of social groups could materialize in opposition towards organizations and institutions that support or give voice to marginalized groups. Such assumptions may be especially likely when people also hold zero-sum beliefs about social groups (e.g., as one group's situation improves that must mean that another group's has worsened; Norton & Sommers, 2011). Another example comes from our own experience seeing this shortcut play out: An author's son took issue with his sister's "Girls are strong" t-shirt and shouted, "Boys aren't weak!"[31] We could also imagine how this form of binary thinking might contribute to the growing political divide in the United States. For example, even before a Democrat learns the contents of a Republican-supported bill, they may assume that if Republicans like the bill, then they (as a Democrat) will not like it. These ideas are currently speculative and will require rigorous experimental testing to confirm or refute them.

Additional aspects of the dichotomizing heuristic should be explored in future research. For example, our paradigm oversimplified social group membership: Individuals were either

[31]Statements like, "Girls are strong" could also be problematic for reasons beyond the dichotomizing heuristic because if people have a stereotype about a group then these sorts of statements might be read as the speaker lacking confidence because people do not state the obvious (see Chestnut & Markman, 2018; Chestnut et al., 2021).

Zuttles or Twiggums. In reality, people belong to multiple social groups. Although we did introduce the presence of multiple groups in Study 3, it also will be important to test how intersectionality influences the dichotomizing heuristic. Indeed, manipulating the saliency of an individual's different social group memberships (see Chiao et al., 2006; Gaither et al., 2014) might alter their use of the dichotomizing heuristic. We also look forward to exploring sources of individual differences in use of the dichotomizing heuristic as well as potential connections between dichotomizing and other outcomes. For example, in Study 3, we found that individuals who assumed greater differences between the two groups also expected them to live in more segregated communities. We imagine that dichotomizing groups may have several other negative downstream consequences. For example, it may be used for justifying social exclusion in an intergroup context. That is, when deciding whom to befriend 7- to 16-year-olds weight shared interests over ethnic group affiliation when provided with both pieces of information (Hitti & Killen, 2015; McGlothlin & Killen, 2005). Because people presume that preferences and abilities do not cross group lines, members of outgroups may be prematurely excluded: Shared preferences cannot be prioritized if they remain undiscovered.

Finally, our contributions to understanding the developmental trajectory of the dichotomizing heuristic are limited. We found that children younger than 8 did not assume that groups have opposite preferences and abilities. We primarily used these developmental findings to bolster the argument that people must perceive of the groups as separate before they will dichotomize their characteristics (i.e., younger children form weaker category boundaries). We discontinued our investigation of children's use of the heuristic in Studies 2 and 3 given the need to more strongly establish the triggers, scope, and boundaries of the phenomenon in adults. We direct interested readers to Moty and Rhodes (2021) who showed that when groups are portrayed

as saliently distinct (e.g., they have different category labels and can be distinguished by their clothes) children as young as 4.5 years will assume between-group differences. Still, from that work it is unclear how far children will generalize the heuristic (i.e., the characteristics they used were all benign—would young children also generalize to novel, evaluative, or universal traits?). Further inquiry into the development of this cognitive shortcut is needed.

Conclusion

Individuals apply a dichotomizing heuristic when learning about social groups: They assume that what is true of one group is not true of another. In addition to documenting this cognitive shortcut, we also elucidated the circumstances under which it arises. Whereas conceptualizing the two groups as distinct is a necessary trigger of the dichotomizing heuristic and generic language heightens its use, intergroup conflict, learning about only two groups, and binary (versus scaled) judgments are not required. Once activated, adults apply this rule-of-thumb broadly, even to evaluative traits, such as morals, intelligence, and work ethic; refraining only when dichotomizing would lead them to assign highly unusual traits to the uncharacterized group (e.g., disliking feeling happy). Taken together, individuals can think that they know about two social groups when they only have learned about one.

References

Allport, G.W. (1954). *The nature of prejudice.* Addison-Wesley, Oxford.

Bigler, R.S., (1995). The role of classification skill in moderating environmental influences on children's gender stereotyping: A study of the functional use of gender in the classroom. *Child Development, 66*, 1072-1087. doi: 10.1111/j.1467-8624.1995.tb00923.x

Bigler, R.S., Jones, L.C., & Lobliner, D.B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child Development, 68*, 530-543. doi: 10.1111/j.1467-8624.1997.tb01956.x

Bohn, M., & Frank, M.C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*. doi: 10.1146/annurev-devpsych-121318-0850537

Chambers, C.G., Graham, S.A., & Turner, J.N. (2008). When hearsay trumps evidence: How generic language guides preschoolers' inferences about unfamiliar things. *Language and Cognitive Processes, 23*, 749-766. doi: 10.1080/01690960701786111

Chiao, J.Y., Heck, H.E., Nakayama, K., & Ambady, N. (2006). Priming race in biracial observers affects visual search for black and white faces. *Psychological Science, 17*, 387-392. doi: 10.1111/j.1467-9280.2006.01717.x

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics, 6*, 339-405. doi: 10.1023/A:100832421

Cimpian, A., Brandone, A.C., & Gelman, S.A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science, 34*, 1452-1482. doi: 10.1111/j.1551-6709.2010.01126.x

Cimpian, A., & Scott, R.M. (2012). Children expect generic knowledge to be widely shared. *Cognition, 123*, 419-433. doi: 10.1016/j.cognition.2012.02.003

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18. doi: 10.10137/0022-3514.56.1.5

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology, 37*, 630-641. doi: 10.1037/0012-1649.37.5.630

Diesendruck, G., Salzer, S., Kushnir, T., & Xu, F. (2015). When choices are not personal: The effect of statistical and social cues on children's inferences about the scope of preferences. *Journal of Cognition and Development, 16*, 370-380. doi: 10.1080/15428372.2013.848870

Dovidio, J.F., Gaertner, S.L., Hodson, G., Houlette, M.A., & Johnson, K.M. (2005). Social inclusion and exclusion: Recategorization and the perception of intergroup boundaries. In D. Abrams, M.A. Hogg, & J.M. Marques (Eds.), *The social psychology of inclusion and exclusion* (pp. 245-264). Psychology Press.

Dukes, K.N., & Gaither, S.E. (2017). Black racial stereotypes and victim blaming: Implications for media coverage and criminal proceedings in cases of police violence against racial and ethnic minorities. *Journal of Social Issues, 73*, 789-807. doi: 10.1111/josi.12248

Eiser, J.R. (1971). Enhancement of contrast in absolute judgment of attitude statements. *Journal of Personality and Social Psychology, 17*, 1-10. doi: 10.1037/h0030455

Esses, V.M., Jackson, L.M., Dovidio, J.F., & Hodson, G. (2005). Instrumental relations among groups: Group competition, conflict, and prejudice. In J.F. Dovidio, P. Glick, & L.A. Rudman (Eds.), *On the nature of prejudice* (pp. 227-243). Blackwell Publishing Ltd.

FeldmanHall, O., & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature Human Behavior, 3*, 426-435. doi: 10.1038/s41562-019-0590-x

Gaither, S.E., Chen, E.E., Corriveau, K.H., Harris, P.L., Ambady, N., & Sommers, S.R. (2014).

Monoracial and biracial children: Effects of racial identity saliency on social learning and social preferences. *Child Development, 85,* 2299-2316. doi: 10.1111/cdev.12266

Gelman, S.A. (2003). *The essential child: Origins of essentialism in everyday though.* Oxford University Press.

Gelman, S.A., Goetz, P.J., Sarnecka, B.W., & Flukes, J. (2008). Generic language in parent-child conversations. *Language Learning and Development, 4*, 1-31. doi: 10.1080/15474440701542625

Gelman, S.A, & Markman, E.M. (1986). Categories and induction in young children. *Cognition, 23*, 183-209. doi: 10.1016/0010-0277(86)90034-X

Gelman, S.A., Star, J.R., Flukes, J.E. (2002). Children's use of generics in inductive inferences. *Journal of Cognition and Development, 3*, 179-199. doi: 10.1207/S15327647JCD0302_3

Gelman, S.A., Ware, E.A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology, 61,* 273-301. doi: 10.1016/j.cogpsych.2010.06.001

Goldfarb, D., Lagattuta, K.H., Kramer, H.J., Kennedy, K., & Tashjian, S.M. (2017). When your kind cannot live here: How generic language and criminal sanctions shape social categorization. *Psychological Science, 28*, 1597-1609. doi: 10.1177/0956797617714827

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and Semantics Volume 3*, (pp. 41-58). New York: Academic Press.

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology, 53*, 310-344. doi: 10.1016/j.cogpsych.2006.04.003

Hamilton, D.L., & Gifford, R.K. (1976). Illusory correlation in intergroup perception: A

cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*, 392-407. doi: 10.1016/S0022-1031(76)80006-6

Heyman, G.D., & Gelman, S.A. (2000a). Preschool children's use of novel predicates to make inductive inferences about people. *Cognitive Development, 15*, 263-280. doi: 10.1016/S0885-2014(00)00028-9

Heyman, G.D., & Gelman, S.A. (2000b). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology, 77*, 1-19. doi: 10.1006/jecp.1999.2555

Hitti, A., & Killen, M. (2015). Expectations about ethnic peer group inclusivity: The role of shared interests, group norms, and stereotypes. *Child Development, 86*, 1522-1537. doi: 10.1111/cdev.12393

Hugenberg, K., Young, S.G., Bernstein, M.J., & Sacco, D.F. (2010). The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychological Review, 117*, 1168-1187. doi: 10.1037/a0020463

Jacobs, J.E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development, 62*, 166-178. doi: 10.1111/j.1467-8624.1991.tb01522.x

Johnston, K.E. & Jacobs, J.E. (2003). Children's illusory correlations: The role of attentional bias in group impression formation. *Journal of Cognition and Development, 4*, 129-160. doi: 10.1207/S15327647JCD0402_01

Kahneman, D. (2011) *Thinking fast and slow.* FSG Adult.

Kalish, C. (2012). Generalizing norms and preferences within social categories and individuals. *Developmental Psychology, 48*, 1133-1143. doi: 10.1037/a0026344

Killen, M. (2007). Children's social and moral reasoning about exclusion. *Current Directions in Psychological Science, 16,* 32-36. doi: 10.1111/j.1467-8721.2007.00470.x

Krueger, J., & Clement, R.W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology, 67,* 596-610. doi: 10.1037/0022-3514.67.4.596

Krueger, J., & Rothbart, M. (1990). Contrast and accentuation effects in category learning. *Journal of Personality and Social Psychology, 59,* 651-663. doi: 10.1037/0022-3514.59.4.651

Lagattuta, K.H. (2008). Young children's knowledge about the influence of thoughts on emotions in rule situations. *Developmental Science, 11,* 809-818. doi: 10.1111/j.1467-7687.2008.00727.x

Lagattuta, K.H., & Kramer, H.J. (2021). Advanced emotion understanding: Children's and adults' knowledge that minds generalize from prior emotional events. *Emotion, 21,* 1-16. doi: 10.1037/emo0000694

Lagattuta, K.H., & Sayfan, L. (2013). Not all past events are equal: Biased attention and emerging heuristics in children's past-to-future forecasting. *Child Development, 84,* 2094-2111. doi: 10.1111/cdev.12082

Lagattuta, K.H., Tashjian, S.M., & Kramer, H.J. (2018). Does the past shape anticipation for the future? Contributions of age and executive function to advanced theory of mind. *Zeitschrift für Psychologie, 226,* 122-133. doi: 10.1027/2151-2604/a000328

Lara, K.H., Lagattuta, K.H., & Kramer, H.J. (2019). Is there a downside to anticipating the upside? Children's and adults' reasoning about how prior expectations shape future emotions. *Child Development, 90,* 1170-1184. doi: 10.111/cdev.12994

Lawson, C.A., & Bower, C.A. (2014). Illusory correlations in preschoolers. *Cognitive Development, 31,* 22-34. doi : 10.1016/j.cogdev.2014.02.005

Leslie, S.J. (2008). Generics: cognition and acquisition. *Philosophical Review, 117,* 1-47. doi: 10.1215/00318108-2007-023

Markman, E.M., Wasow, J.L., & Hansen, M.B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology, 47,* 241-275. doi: 10.1016/S0010-0285(03)00034-3

Master, A., Markman, E.M., & Dweck, C.S. (2012). Thinking in categories or along a continuum: Consequences for children's social judgments. *Child Development, 83,* 1145-63. doi: 10.1111/j.1467-8624.2012.01774.x

McGarty, C., & Penny, R.E.C. (1988). Categorization, accentuation and social judgment. *British Journal of Social Psychology, 27,* 147-157. doi: 10.1111/j.2044-8309.1988.tb00813.x

McGarty, C., & Turner, J.C. (1992). The effects of categorization on social judgment. *British Journal of Social Psychology, 31,* 253-268. doi: 10.111/j.2044-8308.1992.tb00871.x

McGlothlin, H., & Killen, M. (2005). Children's perceptions of intergroup and intragroup similarity and the role of social experience. *Journal of Applied Developmental Psychology, 26,* 680-698. doi: 10.1016/j.appdev.2005.08.008

Moty, K., & Rhodes, M. (2021). The unintended consequences of the things we say: What generics communicate to children about unmentioned categories. *Psychological Science, 32,* 189-203. doi: 10.31234/osf.io/zkjyr

Norton, M. I., & Sommers, S. R. (2011). Whites see racism as a zero-sum game that they are losing. *Perspectives in Psychological Science, 6,* 215-218. doi: 10.1177/1745691611406922

Papfragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics—pragmatics interface. *Cognition, 86*, 253-282. doi: 10.1016/S0010-0277(02)00179-8

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology 42*, 1051-1068. doi: 10.1037/0022-3514.42.6.1051

Qian, M.K., Quinn, P.C., Heyman, G.D., Pascalis, O., Fu, G., & Lee, K. (2019). A long-term effect of perceptual individuation training on reducing implicit racial bias in preschool children. *Child Development, 90,* e290-e305. doi: 10.1111/cdev.12971

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for statistical computing Vienna, Austria. URL https://www.R-project.org/

Reyna, V.F., & Ellis, S.C. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science, 5*, 275-279. doi: 10.1111/j.1467-9280.1994.tb00625.x

Rhodes, M., & Gelman, S.A. (2008). Categories influence predictions about individual consistency. *Child Development, 79*, 1270-1287. doi : 10.1111/j.1467-8624.2008.01188.x

Rhodes, M., Leslie, S.J., & Tworek, C.M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences, USA, 109*, 13526-13531. doi: 10.1073/pnas.1208951109

Richter, N., Over, H., & Dunham, Y. (2016). The effects of minimal group membership on young preschoolers' social preferences, estimates of similarity, and behavioral attribution. *Collabra, 2*, 8. doi: 10.1525/collabra.44

Riggs, A.E., Kalish, C.W., & Alibali, M.W. (2014). When you've seen one, have you seen them

all? Children's memory for general and specific learning episodes *Developmental Psychology, 50,* 1653-1659. doi: 10.1037/a0036130

Roberts, S.O., Gelman, S.A., & Ho, A.K. (2017). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science, 41,* 576-600. doi: 10.1111/cogs.12443

Roberts, S.O., Guo, C., Ho, A.K., & Gelman, S.A. (2018). Children's descriptive-to-prescriptive tendency replicates (and varies) cross-culturally: Evidence from China. *Journal of Experimental Child Psychology, 165,* 148-160. doi: 10.1016/j.jecp.2017.03.018

Rothbart, M., Fulero, S., Jensen, C., Howard, J., & Birrell, P. (1978). From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology, 14,* 237-255. doi: 10.1016/0022-1031(78)90013-6

Schwarz, N., & Bless, H. (2007). Mental construal process: The inclusion/exclusion model. In D.A. Stapel & J. Suls (Eds.), *Assimilation and contrast in social psychology* (pp. 119-141). Philadelphia, PA: Psychology Press.

Sherman, S.J., & Corty (1984). Cognitive heuristics. In R.S. Wyer, J.R. & T.K. Srull (Eds.). *Handbook of social cognition Vol. 1* (pp. 189-286). Lawrence Erlbaum Associates Publishers.

Sherman, J.W., Kruschke, J.K., Sherman, S.J., Percy, E.J., Petrocelli, J.V., & Conrey, F.R. (2009). Attentional processing in stereotype formation: A common model for category accentuation and illusory correlation. *Journal of Personality and Social Psychology, 96,* 305-323. doi: 10.1037/a0013778

Shutts, K. (2015). Young children's preferences: Gender, race, and social status. *Child Development Perspectives, 9,* 262-266. doi: 10.1111/cdep.12154

Shutts, K., Pemberton, C.K., & Spelke, E.S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development, 14*, 35-62. doi: 10.1080/15248372.2011.638686

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluative of replication results. *Psychological Science, 26,* 559-569. doi: 10.1177/0956797614567341

Sloutsky, V.M., Deng, W.S., Fisher, A.V., & Koos, H. (2015). Conceptual influences on induction: A case for late onset. *Cognitive Psychology, 82*, 1-31. doi: 10.1016/j.cogpsych.2015.08.005

Sunstein, C.R. (2005). Moral heuristics. *Behavioral Brain Sciences, 28,* 531-542. doi: 10.1017/S0140525X05000099

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology, 33*, 1-39. doi: 10.1146/annurev.ps.33.020182.000245

Tajfel, H., & Wilkes, A.L. (1963). Classification and quantitative judgment. *British Journal of Psychology, 54,* 101-114. doi: 10.1111/j.2044-8295.1963.tb00865.x

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232. doi: 10.1016/0010-0285(73)90033-9

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 776-795. doi: 10.1037//0278-7393.26.3.776

Young, S.G., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of Personality and Social Psychology, 99*, 964-977. doi: 10.1037/a0020400

Appendix A

Chapter 1: This Too Shall Pass, But When? Children's and Adults' Beliefs about the Duration of Emotions, Desires, and Preferences

Table A.1. Means and 95% Confidence Intervals of Duration Estimates and Certainty by Item

State	Duration	Certainty
Spencer does not want milk.	4.64 [4.02, 5.27]	56.80 [48.11, 65.49]
Nicky does not like milk.	8.29 [7.95, 8.63]	74.44 [67.63, 81.26]
Ari has a goal to be a doctor.	7.73 [7.47, 7.99]	73.44 [67.26, 79.62]
Sidney does not have a goal to be a doctor.	8.40 [8.02, 8.78]	73.13 [66.22, 80.05]
Cam feels thirsty.	2.38 [2.22, 2.54]	63.02 [54.16, 71.89]
Logan does not feel thirsty.	2.82 [2.69, 2.95]	76.98 [70.85, 83.10]
Jamie feels hungry.	2.69 [2.36, 3.01]	64.00 [55.52, 72.48]
Tatum does not feel hungry.	2.96 [2.89, 3.02]	75.29 [68.33, 82.25]
Piper knows that cats meow.	8.93 [8.83, 9.03]	91.96 [87.15, 96.76]
Tyler does not know that cats meow.	5.40 [4.65, 6.15]	56.87 [47.16, 66.58]
Adrian knows the names of 43 cat breeds.	7.67 [7.25, 8.09]	72.98 [65.46, 80.50]
Skyler does not know the names of 43 cat breeds.	6.98 [6.36, 7.60]	59.84 [50.82, 68.87]
Andy can draw a triangle.	8.47 [7.96, 8.97]	90.27 [85.66, 94.88]
Sawyer cannot draw a triangle.	5.20 [4.36, 6.04]	58.87 [50.54, 67.19]
Charlie can draw realistic horses.	8.36 [8.00, 8.71]	80.24 [73.23, 87.26]
Bailey cannot draw realistic horses.	7.02 [6.58, 7.47]	69.44 [51.13, 67.76]
Tanner and Kyle are friends.	7.64 [7.37, 7.92]	62.20 [54.07, 70.33]
Austin and Riley are enemies.	6.84 [6.42, 7.27]	54.16 [45.81, 62.50]
Harper is mean.	7.18 [6.33, 7.72]	53.53 [44.87, 62.19]
Taylor is nice.	7.89 [7.33, 8.45]	71.33 [64.85, 77.82]
Ryan is male.	8.96 [8.89, 9.02]	86.71 [79.40, 94.02]
Reagan is female.	8.84 [8.61, 9.07]	84.47 [77.89, 91.04]
Sam is black.	8.89 [8.66, 9.11]	96.47 [93.81, 99.13]
Reese is white.	9.00 [no variability]	97.24 [94.92, 99.57]

Appendix B

Chapter 2: Consistency Among Social Groups in Judging Emotions Across Time

Table B.1. Additional demographic information for Study 1 recruited participants

Gender		
	Cisgender female	50.0%
	Cisgender male	46.0%
	Genderqueer, gender non-conforming, gender non-binary	4.0%
Occupation		
	Employed	16.5%
	Unemployed	6.7%
	Unable to work due to disability	0.4%
	Homemaker	0.8%
	Employed student	37.0%
	Unemployed student	38.6%
Education		
	High school diploma	11.8%
	Vocational training certificate	0.8%
	Completed some university courses	50.4%
	College degree	37.0%
Income		
	< $20K	80.3%
	< $30K	7.9%
	< $40K	2.8%
	< $50K	2.0%
	< $50K-$200K+	7.0%

Note. The reported percentages represent characteristics of the 254 participants in Study 1.

Table B.2. Means and Standard Deviation of Events in Isolation by Study and Event

Someone…	Study 1	Study 2
Beats you up	-2.80 (0.69)	
Threatened you with a weapon	-2.86 (0.49)	
Falsely accused you.	-2.54 (0.88)	
Ignored your pleads for help.	-2.43 (0.84)	-1.97 (1.04)
Rejected you.	-2.12 (0.92)	-1.59 (0.98)
Disrespected your point of view.	-1.50 (0.99)	
Excluded you.	-2.06 (0.93)	-1.51 (1.08)
Called you a derogatory word.	-2.00 (1.05)	-2.08 (0.79)
Bullied you.	-2.46 (0.83)	-2.21 (0.74)
Damaged your property.	-2.39 (0.80)	-2.51 (0.71)
Intimidated you.	-1.68 (0.93)	
Made fun of you.	-1.82 (1.14)	-1.92 (0.84)
Threw something at you.	-2.00 (0.97)	-2.32 (0.75)
Rejected your invitation.	-1.15 (0.89)	
Didn't listen to your problems.	-1.59 (0.98)	
Avoided sitting next to you.	-1.37 (1.06)	
Mispronounced your name.	-0.32 (0.63)	
Criticized you.	-1.28 (1.09)	
Forgot your name.	-0.58 (0.76)	
Hugged you.	2.09 (1.05)	
Invited you to meet family.	1.85 (1.06)	
Bought you snacks.	2.16 (0.87)	
Tried to cheer you up.	2.00 (0.90)	1.91 (0.92)
Loaned you money.	0.82 (1.41)	
Praised you.	2.26 (0.92)	2.03 (0.94)
Included you.	2.31 (0.81)	
Accepted your invitation.	2.04 (0.93)	1.79 (0.93)
Offered you a seat.	1.68 (0.94)	
Helped you complete a task.	1.98 (0.91)	2.21 (0.79)
Listened to you.	2.29 (0.84)	2.04 (0.80)
Respected you.	2.54 (0.74)	
Gave you a desirable gift.	2.42 (0.74)	2.10 (0.93)
Celebrated your accomplishments.	2.44 (0.85)	2.10 (0.90)
Stood up for you.	2.35 (1.31)	2.53 (0.79)
Gave you something you need.	2.35 (0.74)	
Recommended you for a job.	2.56 (0.72)	

Note. Valence Ratings: -3 = Very Negative; -2 = Medium Negative; -1 = A Little Negative; 0 = Neutral (Not Negative or Positive); 1 = A Little Positive; 2 = Medium Positive; 3 = Very Positive; Items that are bold face were included in Study 2.

Table B.3. Additional demographic information for Study 2 participants

Gender		
	Cisgender female	48.1%
	Cisgender male	39.2%
	Genderqueer or gender non-conforming	2.5%
	Transgender	8.2%
	Other gender*	2.0%
Occupation		
	Employed	50.0%
	Unemployed	6.3%
	On non-medical leave of absence	1.3%
	Employed student	27.2%
	Unemployed student	15.2%
Education		
	High school diploma	2.5%
	Vocational training certificate	2.5%
	Completed some university courses	14.0%
	College degree	81.0%
Income		
	< $20K	48.7%
	< $30K	22.2%
	< $40K	10.1%
	< $50K	10.1%
	> $50K <$90K	8.9%

Note: Participants who selected "Other gender," did not specify their gender via the open-ended response option.

Table B.4. Study 2: *Means and Standard Deviation of Events in Sequence by Event*

Someone…	NN Initial	NN Recent	NP Initial	PN Recent
Ignored your pleads for help.	-1.83 (0.74)	-2.11 (0.83)	-1.92 (0.84)	-2.04 (1.02)
Rejected you.	-1.75 (0.84)	-2.14 (0.89)	-1.45 (1.03)	-1.52 (0.99)
Excluded you.	-1.53 (1.09)	-2.15 (0.84)	-1.56 (0.62)	-1.76 (0.69)
Called you a derogatory word.	-2.04 (0.90)	-2.57 (0.62)	-2.13 (0.84)	-2.18 (0.82)
Bullied you.	-2.34 (0.80)	-2.59 (0.62)	-2.14 (0.71)	-2.28 (0.72)
Damaged your property.	-2.44 (0.73)	-2.89 (0.42)	-2.53 (0.68)	-2.30 (0.82)
Made fun of you.	-2.03 (0.74)	-2.60 (0.62)	-1.66 (0.88)	-1.64 (0.87)
Threw something at you.	-2.20 (0.72)	-2.63 (0.59)	-2.14 (0.77)	-2.15 (1.08)
Someone…	PP Initial	PP Recent	PN Initial	NP Recent
Tried to cheer you up.	1.88 (0.73)	2.19 (0.85)	1.77 (0.69)	1.15 (1.29)
Praised you.	1.68 (0.96)	2.09 (0.96)	2.14 (0.75)	0.29 (1.58)
Accepted your invitation.	1.57 (0.65)	2.12 (0.68)	1.77 (0.93)	0.27 (1.44)
Helped you complete a task.	2.04 (0.68)	1.93 (1.07)	1.95 (1.00)	0.80 (1.19)
Listened to you.	1.83 (0.76)	2.48 (0.70)	1.74 (0.83)	0.63 (1.30)
Gave you a desirable gift.	2.29 (0.87)	2.14 (0.75)	2.30 (0.73)	0.17 (1.48)
Celebrated your accomplishments.	2.23 (0.81)	2.38 (0.70)	2.00 (0.96)	0.39 (1.34)
Stood up for you.	2.45 (0.69)	2.46 (0.90)	2.52 (0.93)	1.08 (1.33)

Note. Valence Ratings: -3 = Very Bad; -2 = Medium Bad; -1 = A Little Bad; = A Little Bad; 2 = Medium Bad; 3 = Very Bad

Appendix C

Chapter 3: Dichotomous Thinking about Social Groups: Learning about One Group Can Activate Opposite Beliefs about Another Group

Study 2 Pilot Study

The aim of this pilot study for Study 2 was to conduct a preliminary test of whether the dichotomizing heuristic is still present if participants do not first complete the social categorization task. A secondary goal was to assess whether participants would exhibit the dichotomizing heuristic when they have less information about the characterized group. To that end, we manipulated whether participants learned about 12 Zuttle characteristics (*standard familiarization*) or only 8 Zuttle characteristics (*brief familiarization*). Finally, we examined participants' decision confidence when completing the uncharacterized group judgment task.

Method

Participants. *A priori,* we set a sample size of 53 participants per cell (N = 106) based on Simonsohn's (2015) recommendation of multiplying the original sample (n = 21 adults in the generic language condition) by 2.5. Inclusion criteria included being between the ages of 18 and 24, not having any chronic psychological problems, and being fluent in English. Although we prescreened based on inclusion criteria, some participants initially stated that they were within the correct age range, but then reported an excluded age during participation. For this reason, our initial sample was larger than we originally planned (N = 166). After excluding participants who did not fit our intended age range, our sample was (N = 103); we also excluded four additional participants who failed at least one attention check (final N = 99). Of the final participant sample, 3% were American Indian, 6% were Asian, 12% were Black or African American, 3% were Hispanic/Latino, 69% were White, and 7% were multiracial, multiethnic, or reported "other".

Participants were randomly assigned to one of two between-subjects familiarization conditions: brief familiarization (n = 50; M = 21.98 years, SD = 1.46 years, 32 males, 17 females, 1 self-described as "Agender") and standard familiarization (n = 49, M = 22.76 years, SD = 1.45 years, 27 males, 22 females). Data were collected in July 2018 on Mechanical Turk using TurkPrime (Litman, Robinson, & Abberbock, 2017). Participants received compensation for their participation. This study was approved by the Internal Review Board at [BLINDED]: Protocol Number: 1031991. The sample size, method, and planned analyses were pre-registered on Open Science Framework (https://osf.io/jhn3e/?view_only=3a4f8b2f2b20406aaf34ae175fe43629).[32]

Procedure. Participants read the following information:

> "Please imagine that there is a new group of humans called Twiggums. Twiggums live on an island called Twiggum Island. One day, some Twiggums leave Twiggum Island by boat to live on another island. On this new island, this group of Twiggums renamed themselves Zuttles. So, where did the first group of Zuttles move away from? (choices: Twiggum Island; Zuttle Island). Then, the Twiggums made a law that said Zuttles CANNOT live on Twiggum Island. So, are Zuttles allowed to live on Twiggum Island (choices: Yes, Zuttles are allowed to live on Twiggum Island; No, Zuttles are not allowed to live on Twiggum Island). Click the "Next" button to learn about the Zuttles."

Next, participants learned about the Zuttles' preferences in generic language, describing the group as a general whole (e.g., Zuttles like pears) and abilities (e.g., Zuttles are good at play

[32]Note that pre-registration for this study is labeled Adults' use of a dichotomizing heuristic (Study 2) on OSF.

drums). This information was displayed pictorially as well as in written format. Half of the participants learned about eight characteristics (*brief familiarization condition*) and half of the participants learned about twelve characteristics (*standard familiarization condition*).

After this familiarization, participants did not complete the social categorization task (as in Study 1). Instead, they immediately responded to the *uncharacterized group judgment task*. They answered questions about Twiggums (eight trials; e.g., So, what do you think about Twiggums? Do you think that Twiggums like pears or do not like pears?). They also reported how confident they were about each decision (100-point scale: Very Unsure to Very Sure). While making decisions, participants had access to the all of the information that they learned about the characteristics of Zuttles. This information appeared in one of 24 blocked orders.[33]

Results

Analyses were conducted in RStudio (R Core Team, 2017). Preliminary analyses verified that the eight characteristics showed excellent internal consistency for both judgments (Chronbach's alpha = .91, CI [.89, .94]) and confidence (Chronbach's alpha = .95, CI [.94, .97]) on the uncharacterized group judgment task. Therefore, we created a proportion of trials dichotomized score as well as an average certainty score for primary analyses.

Participants in both conditions failed to dichotomize the groups. Participants in the brief familiarization condition dichotomized the two groups significantly below chance (.50; overall: M = .38, CI[.26, .49], t[49] = -2.28, p = .027, d = 0.32), participants in the standard familiarization condition did so marginally below chance (overall: M = .41, CI[.29, .52], t[48] = -1.71, p = .094, d = 0.24). We documented similar patterns when we looked exclusively at skill trials (brief: M = .36, CI[.23, .48], t[49] = -2.41, p = .020, d = 0.34; standard: M = .41, CI[.30,

[33]In the pre-registration for this study, we wrote that this information would be presented in one of 96 blocks.

.53], t[48] = -1.52, p = .136, d = 0.22) and exclusively at preference trials (brief: M = .40, CI[.28, .51], t[49] = -1.90, p = .064, d = 0.27; standard: M = .40, CI[.28, .52], t[48] = -1.72, p = .091, d = 0.25). Participants in the two familiarization conditions did not differ in their dichotomization rates (t[96.973] = -0.39, p = .695, d = 0.08).

Participants in both conditions were confident in their judgments (compared to the midpoint of the scale, 50; brief: M = 71.75, CI[67.01, 76.48], t[49] = 9.23, p < .001, d = 1.31, standard: M = 73.46, CI[68.34, 78.59], t[48] = 9.21, p < .001, d = 1.32), and they exhibited comparable levels of confidence (t[96.238] = -0.49, p = .623, d = 0.10). We then examined the relation between confidence and dichotomization rates. We tested linear and quadratic models. In the first step (F[2, 96] = 1.35, p = .264, adjusted R^2 = .01), we entered familiarization condition (b = .04, SE = .08, t = 0.48, p = .636) and participants' average confidence (b = -.003, SE = .002, t = -1.59, p = .114). In the second step (F[3, 95] = 1.02, p = .388, adjusted R^2 < .01), we added the participants average certainty[2] (b < .001, SE < .001, t = -0.61, p = .543). Thus, the relation between dichotomizing and confidence was null in this study.

In addition, we examined how confident participants were depending on their decision (i.e., whether they dichotomized or aligned the groups). Some participants never dichotomized the groups for any trial, and some participants never aligned the groups for any trial. When this happened, these participants were excluded from the relevant analysis. Participants were confident when they dichotomized (compared to the midpoint of the scale [50]; brief [n = 30]: M = 70.55, CI[63.43, 77.67], t[29] = 5.90, p < .001, d = 1.08; standard [n = 31]: M = 68.49, CI[62.01, 74.97], t[30] = 5.83, p < .001, d = 1.05), as well as when they aligned the groups (brief [n = 42]: M = 72.28, CI[67.12, 77.44], t[41] = 8.72, p < .001, d = 1.34; standard [n = 40]: M = 74.51, CI[68.52, 80.52], t[39] = 8.26, p < .001, d = 1.31).

Finally, we examined the individual decision patterns in the data: 59% (brief: 62%; standard: 55%) dichotomized the groups on less than 50% of trials; 9% (brief: 8%; standard: 10%) of participants did so on 50% of trials, and 32% (brief: 30%; standard: 35%) of participants did so on more than 50% of trials.

Discussion

The results of this study provided initial evidence that completing the social categorization task prior to the uncharacterized group task may be necessary for participants to exhibit the dichotomizing heuristic (though, it is notable that about 30% of participants dichotomized the groups on more than 50% of trials even when there was no categorization task). Importantly, however, we did not systematically manipulate whether or not participants completed the social categorization task in this study. Thus, we provide a stricter test of the necessity of the social categorization task in Study 2 using a between-subjects design (main text). Although there were no familiarization condition differences in this study (i.e., brief familiarization = standard familiarization in proportion of trials dichotomized), we used the standard 12-item familiarization in Study 2 (main text) to mirror Study1's design (main text). Moreover, we could not draw conclusions from this study about the type of familiarization required because neither condition elicited dichotomizing rates above chance. Finally, irrespective of their dichotomizing decisions, participants appeared confident in their judgments, both when they assumed that the groups would differ in their characteristics and when they expected commonalties. We removed confidence ratings from Study 2 (main text) because these questions did not add useful information and increased the length of the experimental session.

Study 2 Additional Analyses

Performance on the Social Categorization Task

It was not main a goal of this study to replicate the findings from Goldfarb et al. (2017; i.e., the following analyses were not pre-registered). We were, however, able to explore whether certain effects from that study replicate (i.e., we cannot examine age or language effects because these variables were not included in this study). In particular, we could address two of the central effects from Goldfarb et al. (2017). First, within the generic language condition, is adults' willingness to categorize unknown individuals as Zuttles dependent upon the consequences of that decision (i.e., being deported vs. being invited to a "welcome" party)? In Goldfarb et al. (2017), participants who learned about Zuttles in generic language were just as likely to categorize unknown creatures as Zuttles when there were consequences (i.e., being sent to jail and then deported) as when there were no consequences. Second, is adults' willingness to categorize individuals dependent upon the evidence available? In Goldfarb et al. (2017), adults more frequently categorized individuals as Zuttles when more evidence exactly matched what participants had previously learned about Zuttles. Adults in the generic language condition also categorized individuals as Zuttles at above chance levels when one or more pieces of evidence (out of three possible) matched what they had learned about Zuttle characteristics.

To examine these effects, we first conducted a 2 (stance: negative, positive) x 4 (evidence: zero, one, two, three) repeated measures ANOVA on the proportion of trials that participants decided an unknown creature was a Zuttle. This analysis resulted in only a main effect for evidence, $F(3, 306) = 415.48$, $p < .001$, $\eta_p^2 = .80$. There were no effects for stance, Fs < 0.44, ps $> .510$, $\eta_p^2 < .01$. Thus, conceptually replicating Goldfarb et al. (2017), participants categorized individuals as Zuttles regardless of if the consequence was deportation or being

invited to a “welcome” party. Also replicating Goldfarb et al. (2017), we found that participants were more likely to categorize individuals as Zuttles the more evidence there was that matched what they learned about Zuttles: Three Exact, Two Exact > One Exact > Zero Exact (*p*s <.001). Furthermore, adults categorized individuals as Zuttles at above chance levels when one or more pieces of evidence matched what they had previously learned about Zuttles (*t*[103]s > 12.36, *p*s < .001, *d*s > 1.21; One Exact: *M* = .80, 95% CI[0.75, 0.85]; Two Exact: *M* = 0.94, 95% CI [.90, 0.98]; Three Exact: *M* = 0.97, 95% CI[0.94, 1.00]). They were less likely than expected by chance to categorize individuals as Zuttles when no evidence matched what they had learned about Zuttles (*t*[103] = -8.90, *p* < .001, *d* = 0.87; *M* = .34, 95% CI[.30, .37]).

Correlations Between the Social Categorization Task and the Uncharacterized Group Judgment Task

As part of our pre-registered analysis plan, we examined relations between the social categorization task and proportion of trials dichotomized. Note that these analyses were only conducted with participants who completed the social categorization task (*n* = 104). We set our alpha level at *p* < .004 (.05/12 tests). Participants who more frequently categorized unknown individuals as Zuttles when two or three features exactly matched what they had learned about Zuttles in the social categorization task more often predicted in the uncharacterized group judgment task that Twiggums (uncharacterized group) must have opposite preferences and skills in the initial questioning set (*three exact matches*: *r*[102] = .43, *p* < .001; *two exact matches:* *r*[102] = .34, *p* < .001). All other correlations were not significant, including connections between responses on the social categorization task and the questioning sets for novel and evaluative traits in uncharacterized group judgment task (.01 < *r*s < .24, .917 > *p*s > .016).

Study 3 Additional Analyses

We explored whether there were any differences in the categorization task when participants initially learned that two groups lived on the island versus when they learned that many groups lived on the island. We conducted a 2 (size: two, many) x 4 (evidence: zero, one, two, three) repeated measures ANOVA on the proportion of trials that participants decided an unknown creature was a Zuttle. This analysis resulted in only a main effect for size, $F(1, 101) = 4.24$, $p = .042$, $\eta_p^2 = .04$, and a main effect for evidence, $F(3, 303) = 196.14$, $p < .001$, $\eta_p^2 = .66$. Participants who learned about two groups ($M = .71$, 95% CI [.66, .76]) more often categorized unknown individuals as Zuttles than did those in the many groups condition ($M = .64$, 95% CI [.59, .69], $p = .042$). Also replicating Goldfarb et al. (2017), we found that participants were more likely to categorize individuals as Zuttles the more evidence there was that matched what they learned about Zuttles: Three Exact > Two Exact > One Exact > Zero Exact (*p*s <.001). Furthermore, adults categorized individuals as Zuttles at above chance levels when one or more pieces of evidence matched what they had previously learned about Zuttles (*t*[102]s > 3.83, *p*s < .001, *d*s > 0.38; One Exact: $M = .62$, 95% CI[.56, .69]; Two Exact: $M = 0.83$, 95% CI [.78, .89]; Three Exact: $M = .95$, 95% CI[.92, .98]). They were less likely than expected by chance to categorize individuals as Zuttles when no evidence matched what they had learned about Zuttles (*t*[102] = -9.82, $p < .001$, $d = 0.97$; $M = .30$, 95% CI[.25, .34]).

We also examined relations between the social categorization task and dichotomizing ratings. Note that these analyses were only conducted with participants who completed the social categorization task ($n = 103$). We set our alpha level at $p < .003$ (.05/16 tests). No correlations were significant (.002 < |*r*|s < .22, .986 > *p*s > .025).

Table C.1. Study 1: Percentage of participants who dichotomized Twiggums (uncharacterized group) and Zuttles (characterized group) by Characteristic, Age Group, and Language Condition.

	5-year-olds		8-year-olds		Adults	
	Specific (*n* = 35)	Generic (*n* = 31)	Specific (*n* = 30)	Generic (*n* = 31)	Specific (*n* = 32)	Generic (*n* = 21)
Zuttles like oranges	57	58	47	74	56	76
Zuttles like pears	57	35	57	71	56	76
Zuttles hate chocolate	69	74	73	81	78	95
Zuttles hate candy	49	71	63	87	84	95
Zuttles are good at playing drums	47 (n = 34)	42	53	87	75	100
Zuttles are good at playing piano	54	61	63	81	72	95
Zuttles are bad at playing soccer	54	61	63	87	81	95
Zuttles are bad at playing basketball	57	61	57	68	81	90
Dichotomized >50% of trials	43	55	47	81	69	100

Numbers above the dashed line display the percentage of participants who judged that Twiggums would have opposite preferences and abilities from Zuttles by characteristic type (e.g., 57% of 5-year-olds in the specific-language condition inferred that Twiggums hate oranges). Below the dashed line displays the percentage of individuals who Dichotomized > 50% of trials (i.e., primarily judged that the two groups would have opposite preferences and abilities). For example, 43% of 5-year-olds in the specific-language condition judged that Zuttles and Twiggums would have opposite preferences and abilities on more than 50% of trials.

Table C.2. Study 2: Percentage of Participants Who Dichotomized the Groups on the Initial Set by Characteristic, Categorization, and Stance.

	Categorization-Negative (n = 52)	Categorization-Positive (n = 52)	No-Categorization-Negative (n = 54)	No-Categorization-Positive (n = 52)
Zuttles like oranges	85	81	24	12
Zuttles like pears	83	87	26	10
Zuttles like apples	85	85	26	12
Zuttles hate cake	87	85	28	13
Zuttles hate chocolate	87	88	35	13
Zuttles hate candy	85	85	33	15
Zuttles are good at playing guitar	83	85	28	10
Zuttles are good at playing piano	79	88	28	12
Zuttles are good at playing drums	79	83	26	6
Zuttles are bad at playing basketball	83	85	24	13
Zuttles are bad at playing tennis	81	79	30	10
Zuttles are bad at playing soccer	77	88	26	12
Dichotomized >50% of trials	85	88	24	6

Numbers above the dashed line display the percentage of participants who judged that Twiggums would have opposite preferences and abilities from Zuttles by characteristic type (e.g., 85% of participants in the Cat/Neg condition inferred that Twiggums hate oranges). Below the dashed line displays the percentage of individuals who Dichotomized > 50% of trials (i.e., primarily judged that the two groups would have opposite preferences and abilities). For example, 85% of participants in the Cat/Neg condition judged that Zuttles and Twiggums would have opposite preferences and abilities on more than 50% of trials.

Table C.3. Study 2: Percentage of Participants Who Dichotomized the Groups on the Novel Set by Characteristic, Categorization, and Stance.

	Categorization-Negative (n = 52)	Categorization-Positive (n = 52)	No-Categorization-Negative (n = 54)	No-Categorization-Positive (n = 52)
Zuttles like blickets	73	71	26	8
Zuttles like modies	71	75	26	15
Zuttles like kifs	67	73	30	13
Zuttles hate bupples	77	78 (n = 51)	40 (n = 53)	15
Zuttles hate lonties	73	75 (n = 51)	35	15
Zuttles hate rylos	77	73 (n = 51)	31	13
Zuttles are good at playing lups	73	79	23 (n = 53)	13
Zuttles are good at playing daxes	75	75	30	15
Zuttles are good at playing nafs	73	77	22	13
Zuttles are bad at playing gorps	77	85	36 (n = 53)	10
Zuttles are bad at playing parms	79	83	34 (n = 53)	19
Zuttles are bad at playing wugs	77	83	34 (n = 53)	13
Dichotomized >50% of trials	73	73	20	6

Numbers above the dashed line display the percentage of participants who judged that Twiggums would have opposite preferences and abilities from Zuttles by characteristic type (e.g., 73% of participants in the Cat/Neg condition inferred that Twiggums hate blickets). Below the dashed line displays the percentage of individuals who Dichotomized > 50% of trials (i.e., primarily judged that the two groups would have opposite preferences and abilities). For example, 73% of participants in the Cat/Neg condition judged that Zuttles and Twiggums would have opposite preferences and abilities on more than 50% of trials.

Table C.4. Study 2: Percentage of Participants Who Dichotomized the Groups on the Evaluative Set by Characteristic, Categorization, and Stance.

	Categorization-Negative (n = 52)	Categorization-Positive (n = 52)	No-Categorization-Negative (n = 54)	No-Categorization-Positive (n = 52)
Zuttles are (not) smart	52	63	26	17
Zuttles are (not) creative	69	65	22	23
Zuttles are (not) honest	62	62	24	21
Zuttles are (not) funny	63	71	22	23
Zuttles are (not) shy	67 (n = 51)	65	22	17
Zuttles are (not) confident	63	62	28	23
Zuttles are (not) superstitious	67	58	24	21
Zuttles are (not) liberal	81	67	43	25
Zuttles are (not) lazy	58	58	19	31
Zuttles are (not) mean	69	62	30	38
Zuttles are (not) messy	65	63	28	17
Zuttles are (not) moody	71	71	20	23
Dichotomized >50% of trials	62	62	13	19

Numbers above the dashed line display the percentage of participants who judged that Twiggums would have opposite evaluative traits from Zuttles by characteristic type (e.g., 52% of participants in the Cat/Neg condition inferred that Twiggums were [not] smart depending on what they learned about Zuttles). Below the dashed line displays the percentage of individuals who Dichotomized > 50% of trials (i.e., primarily judged that the two groups would have opposite evaluative traits). For example, 62% of participants in the Cat/Neg condition judged that Zuttles and Twiggums would have opposite evaluative traits on more than 50% of trials.

Table C.5. Study 3: Percentage of Participants Who Dichotomized the Groups on the Initial Set by Characteristic, Group Size, and Categorization.

	Two-Categorization (n = 53)	Two-No-Categorization (n = 51)	Many-Categorization (n = 50)	Many-No-Categorization (n = 52)
Zuttles like oranges	60	45	46	37
Zuttles like pears	62	45	48	44
Zuttles like apples	53	51	44	38
Zuttles hate cake	72	63	66	60
Zuttles hate chocolate	68	61	64	56
Zuttles hate candy	70	59	52	54
Zuttles are good at playing guitar	68	45	52	44
Zuttles are good at playing piano	70	47	54	52
Zuttles are good at playing drums	70	43	60	42
Zuttles are bad at playing basketball	62	53	68	52
Zuttles are bad at playing tennis	64	73	60	52
Zuttles are bad at playing soccer	72	59	58	63
Dichotomized >40% of trials	75	61	66	63

Numbers above the dashed line display the percentage of participants who judged that Twiggums would have preferences and abilities on the opposite side of the scale as Zuttles (i.e., the percentage of participants who selected a 4 or 5) by characteristic type (e.g., 60% of participants in the Two/Cat condition inferred that Twiggums dislike oranges). Below the dashed line displays the percentage of individuals who Dichotomized > 40% of trials (i.e., primarily judged that the two groups would have opposite preferences and abilities more frequently than would be expected by chance). For example, 75% of participants in the Two/Cat condition judged that Zuttles and Twiggums would have opposite preferences and abilities on more than 40% of trials.

Table C.6. Study 3: Percentage of Participants Who Dichotomized the Groups on the Novel Set by Characteristic, Group Size, and Categorization.

	Two-Categorization (n = 53)	Two-No-Categorization (n = 51)	Many-Categorization (n = 50)	Many-No-Categorization (n = 52)
Zuttles like blickets	62	43	46	42
Zuttles like modies	58	39	50	42
Zuttles like kifs	62	41	54	40
Zuttles hate bupples	70	61	64	42
Zuttles hate lonties	58	55	58	48
Zuttles hate rylos	62	57	52	48
Zuttles are good at playing lups	60	47	44	48
Zuttles are good at playing daxes	60	43	46	50
Zuttles are good at playing nafs	57	37	46	46
Zuttles are bad at playing gorps	64	53	54	58
Zuttles are bad at playing parms	60	51	54	52
Zuttles are bad at playing wugs	64	65	50	54
Dichotomized >40% of trials	68	53	56	62

Numbers above the dashed line display the percentage of participants who judged that Twiggums would have preferences and abilities on the opposite side of the scale as Zuttles (i.e., the percentage of participants who selected a 4 or 5) by characteristic type (e.g., 62% of participants in the Two/Cat condition inferred that Twiggums dislike blickets). Below the dashed line displays the percentage of individuals who Dichotomized > 40% of trials (i.e., judged that the two groups would have opposite preferences and abilities more frequently than would be expected by chance). For example, 68% of participants in the Two/Cat condition judged that Zuttles and Twiggums would have opposite preferences and abilities on more than 40% of trials.

Table C.7. Study 3: Percentage of Participants Who Dichotomized the Groups on the Evaluative Set by Characteristic, Group Size, and Categorization.

	Two-Categorization (n = 53)	Two-No-Categorization (n = 51)	Many-Categorization (n = 50)	Many-No-Categorization (n = 52)
Zuttles are (not) smart	57	41	40	42
Zuttles are (not) creative	58	41	50	52
Zuttles are (not) honest	62	49	52	40
Zuttles are (not) funny	64	35	48	50
Zuttles are (not) shy	64	61	60	56
Zuttles are (not) confident	66	41	42	38
Zuttles are (not) superstitious	68	45	46	44
Zuttles are (not) cautious	64	55	44	62
Zuttles are (not) lazy	64	49	46	52
Zuttles are (not) mean	70	49	46	46
Zuttles are (not) messy	62	55	42	48
Zuttles are (not) moody	68	39	40	54
Dichotomized >40% of trials	75	58	58	62

Numbers above the dashed line display the percentage of participants who judged that Twiggums would have preferences and abilities on the opposite side of the scale as Zuttles (i.e., the percentage of participants who selected a 4 or 5) by characteristic type (e.g., 57% of participants in the Two/Cat condition inferred that Twiggums are [not] smart depending on what they learned about Zuttles). Below the dashed line displays the percentage of individuals who Dichotomized > 40% of trials (i.e., judged that the two groups would have opposite preferences and abilities more frequently than would be expected by chance). For example, 75% of participants in the Two/Cat condition judged that Zuttles and Twiggums would have opposite evaluative traits on more than 40% of trials.

Table C.8. Study 3: Percentage of Participants Who Dichotomized the Groups on the Universal Set by Characteristic, Group Size, and Categorization.

	Two-Categorization (n = 53)	Two-No-Categorization (n = 51)	Many-Categorization (n = 50)	Many-No-Categorization (n = 52)
Zuttles dislike getting sick	32	20	30	15
Zuttles dislike getting hurt	42	18	28	21
Zuttles dislike the sound of nails on a chalkboard	43	31	32	31
Zuttles dislike spiders crawling on them	40	18	28	25
Zuttles dislike when their car breaks down	34	20	26	15
Zuttles dislike getting gum in their hair	38	18	24	27
Zuttles like having fun	36	16	28	19
Zuttles like feeling happy	34	18	25	12
Zuttles like seeing stars in the sky	38	22	28	19
Zuttles like winning money	28	10	26	10
Zuttles like sleeping when they are tired	34	18	28	19
Zuttles like eating when they are hungry	42	14	32	13
Dichotomized >40% of trials	35	22	20	10

Numbers above the dashed line display the percentage of participants who judged that Twiggums would have preferences on the opposite side of the scale as Zuttles (i.e., the percentage of participants who selected a 4 or 5) by characteristic type (e.g., 32% of participants in the Two/Cat condition inferred that Twiggums would like getting sick). Below the dashed line displays the percentage of individuals who Dichotomized > 40% of trials (i.e., judged that the two groups would have opposite preferences and abilities more frequently than would be expected by chance). For example, 35% of participants in the Two/Cat condition judged that Zuttles and Twiggums would have opposite preferences on more than 40% of trials.