Promoting Farsighted Decisions via Episodic Future Thinking: A Meta-Analysis

Sarah A. Rösch[a,b,c], **Davide F. Stramaccia**[a], **and Roland G. Benoit**[a,✉]

[a]Max Planck Research Group: Adaptive Memory, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[b]International Max Planck Research School NeuroCom, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[c]Department of Psychosomatic Medicine and Psychotherapy, University of Leipzig Medical Center, Leipzig, Germany

This preprint has not yet been peer-reviewed.

Episodic future thinking (EFT) denotes our capacity to imagine prospective events. It has been suggested to promote farsighted decisions that entail a trade-off between short-term versus long-term gains. Here, we meta-analyze the evidence for the impact of EFT on such intertemporal choices that have monetary or health-relevant consequences. Across 174 effect sizes from 48 articles, a three-level model yielded a medium-sized effect of g = 0.44, 95% CI [0.33, 0.55]. Notably, this analysis included a substantial number of unpublished experiments, and the effect remained stable following further adjustments for remaining publication bias. We exploited the observed heterogeneity to determine critical core components that moderate the impact of EFT. Specifically, the effect was stronger when the imagined events were positive, more vivid, and related to the delayed choice. We further obtained evidence for the contribution of the episodicity and future-orientedness of EFT. Of note, EFT had a greater impact in samples characterized by choice impulsivity (e.g., in obesity), suggesting that EFT can ameliorate maladaptive decision making. Additional analyses indicated that the effect is unlikely to merely reflect demand characteristics. Together, these results highlight the potential of EFT in promoting long-term goals, a finding that extends from the laboratory to real-life decisions.

episodic future thinking | delay discounting | choice impulsivity | self-control | intertemporal choice

Correspondence: ***rbenoit@cbs.mpg.de***

Introduction

Many of our decisions entail a trade-off between short-term gains and long-term consequences. They range from the trivial (e.g., having a second helping versus leaving room for dessert) to the highly consequential (e.g., splurging on a new car versus saving for retirement). In the laboratory, these intertemporal choices are typically examined by having participants choose between small and sooner rewards (e.g., Euro 5 now) versus larger but more delayed rewards (e.g., Euro 10 in two months). A rich line of research has documented that, everything else being equal, people are more likely to forego the larger reward with increasing delay until its delivery (Amasino et al., 2019; Berns et al., 2007; Green & Myerson, 2004; Peters & Büchel, 2011). That is, we tend to assign a smaller subjective value to the otherwise identical reward when we would have to wait longer for it. This phenomenon is most commonly referred to as delay discounting (DD; Green & Myerson, 2004; Mazur, 1987).

Although DD is a ubiquitous feature of intertemporal choices, individuals vary with the extent to which they discount delayed rewards. Importantly, steeper discounters tend to make poorer health-relevant decisions (Amlung et al., 2019; Peters & Büchel, 2011; Story et al., 2014). For example, exaggerated monetary DD has been associated with detrimental behaviors such as physical inactivity (Story et al., 2014), overeating (Appelhans et al., 2011; Rollins et al., 2010), and substance abuse (Bickel et al., 1999; Kollins, 2003; MacKillop et al., 2011). These behaviors implicitly constitute intertemporal choices between immediately available rewards (i.e., indulging in the moment) versus temporally protracted rewards (i.e., being in good health) (Bulley & Schacter, 2020; Epstein et al., 2003; Hollis-Hansen et al., 2019). DD thus generally characterizes choice impulsivity and may constitute a critical trans-disease process (Amlung et al., 2019; Hamilton et al., 2015).

The link between DD and maladaptive behaviors has stimulated research on interventions that may prove effective in nudging individuals towards more farsighted decisions (Peters & Büchel, 2011; Rung & Madden, 2018). These include mindfulness-based approaches (Hendrickson & Rasmussen, 2013; Morrison et al., 2014), framing (Dshemuchadse et al., 2013; Read et al., 2005) and other contextual effects (Dai & Fishbach, 2013; Green et al., 2005), as well as pharmacological interventions (de Wit, 2002). We here examine the impact of episodic future thinking (EFT): the imagination of episodes that may take place in one's personal future (Atance & O'Neill, 2001; Schacter et al., 2017).

EFT draws on memories of the past that get recombined into simulations of prospective events (Addis, 2020; Schacter et al., 2017). EFT allows humans to not only consider the immediate outcomes of an action but also its extended consequences (Atance & O'Neill, 2001; Baumeister & Masicampo, 2010; Boyer, 2008; Suddendorf & Corballis, 2007). It thereby renders an otherwise intangible future more concrete (Lee et al., 2020; Thorstad & Wolff, 2018) and can guide our intertemporal decisions (Baumeister & Masicampo, 2010; Benoit et al., 2018; Boyer, 2008; Bulley et al., 2016).

The Current Meta-Analysis. Over the last decade since the initial reports (Benoit et al., 2011; Peters & Büchel, 2010), there has been rapidly accumulating evidence that EFT indeed attenuates DD of monetary rewards (Rung & Madden, 2018 for an earlier meta-analysis of a very limited number of studies). It also seems to foster more advantageous health-relevant decisions (e.g., Daniel et al., 2013; Dassen et al., 2016). The current meta-analysis has two objectives.

First, we aim to establish the presence and to gauge the effect size of EFT on intertemporal choices. We thus examine the pertinent literature on the discounting of monetary rewards (with a 5-fold increase in studies compared to Rung & Madden, 2018) and assess the generalizability to health-relevant choices (e.g., smoking, alcohol, food). Importantly, we include a large number of unpublished data sets (16% of the included studies) and use several procedures to detect and adjust for the presence of publication bias.

Second, EFT is a complex capacity and we are still lacking a clear understanding of the core components that render it effective in reducing discounting. The behavioral procedures typically ask participants to imagine specific prospective episodes (i.e., the EFT condition) or to engage in one of various other tasks (i.e., the control condition), before they have to make an intertemporal choice (that either affects their monetary pay off or has possible health-relevant consequences). Across studies, however, there is a great heterogeneity in the nature of both the EFT and control conditions.

We make use of that heterogeneity to determine core features of EFT that lead to more farsighted decisions. We accomplish this by conducting a number of targeted moderator analyses. These analyses focus on the (i) *valence* of the imagined event, its (ii) *vividness*, (iii) *episodicity* and (iv) *future-orientedness*. We also examine the impact of the (v) *content specificity* of the imagined event (i.e., whether the imagination entails the consumption of the delayed reward or its possible consequences). Finally, we assess whether EFT is more effective in people who are characterized by greater (vi) *choice impulsivity* (e.g., individuals with obesity). In the following, we introduce these candidate core components.

Candidate Core Components of Episodic Future Thinking.

(i) Valence. When we are faced with an intertemporal choice, we do not experience the positive emotional impact that a delayed reward would hold (e.g., Rick & Loewenstein, 2008). EFT conveys this affective experience by allowing us to mentally visit that future moment (Benoit et al., 2018; Schacter et al., 2015). This positive experience has been hypothesized to add value to the delayed reward and thus to act as a break on impulsive tendencies (Benoit et al., 2011; Boyer, 2008; Frederick et al., 2002; Loewenstein & Lerner, 2003; Rick & Loewenstein, 2008). In contrast, imagining negative episodes may even have the opposite effect of further devaluing the future (Bulley et al., 2016; Liu et al., 2013).

However, the few studies that have directly examined the effect of valence have provided mixed evidence (Bulley et al., 2019; Calluso et al., 2019; Liu et al., 2013; Zhang et al., 2018). We here used the heterogeneity of the episodes' valence across all studies to systematically examine whether EFT more strongly promotes farsighted decisions when the imagined events are positive rather than neutral or negative.

(ii) Vividness. EFT may have a stronger impact on one's future-oriented decisions when the prospective events are imagined more vividly (Bromberg et al., 2015; Ciaramelli et al., 2019; Peters & Büchel, 2010). Related to the previous point, more vividly imagined events elicit stronger affective responses (D'Argembeau & Van der Linden, 2012; Holmes & Mathews, 2010). In addition, vividly imagined events are also deemed more likely or plausible to actually occur (Gregory et al., 1982; Sherman et al., 1985; Szpunar & Schacter, 2013; J. Q. Wu et al., 2015). Vivid imaginings might thus partially attenuate discounting by rendering the future more certain (Bulley et al., 2016). To date, there is only limited evidence that participants who imagine future episodes more vividly are also more likely to choose delayed rewards (Bromberg et al., 2015; Ciaramelli et al., 2019; Peters & Büchel, 2010). To gauge *vividness* as a potential moderator, we focused on the studies that have reported such ratings for both the EFT and control condition. Specifically, we determined the effect sizes of the difference between conditions and included them as a moderator.

(iii) Episodicity. Though intertemporal choices inherently entail a trade-off with the future, there is evidence that people who are apparently stuck in the present do not necessarily show atypical patterns of DD. Specifically, amnesic patients with damage to the medial temporal lobes are also severely impaired in simulating future episodes (McCormick et al., 2018; Race et al., 2011). Yet, *per se*, they do not discount future rewards more strongly than neurotypical persons (De Luca et al., 2018; Kwan et al., 2012, 2013). Critically though, unlike neurotypicals, amnesic individuals do not show reduced delay discounting when given the opportunity to engage in EFT before making their choices (Kwan et al., 2015; Lebreton et al., 2013; Palombo et al., 2015).

The neuropsychological data thus indicate that the ability to construct *episodic* imaginings of the future contributes to the impact of EFT. To further test this hypothesis regarding the episodicity of future thoughts, we specifically examined the studies that compared EFT to a non-episodic, yet also future-oriented control condition (i.e., semantic future thinking; Atance & O'Neill, 2001; Suddendorf & Corballis, 2007).

(iv) Future-Orientedness. EFT makes prospective events more salient and thus changes people's time perspective further into the future (Kurth-Nelson et al., 2012; Lin & Epstein, 2014). This shift towards a greater future-orientedness may lead to more farsighted decisions (Sheffer et al., 2016; Shevorykin et al., 2019).

Supporting this account, there is some general evidence that more future-oriented individuals discount monetary rewards less steeply (Daugherty & Brase, 2010; Joireman et al., 2008; Steinberg et al., 2009) and that they make better health-relevant choices (e.g., Hall et al., 2015).

There is also more specific evidence that episodic *future* thinking reduces discounting and promotes health-relevant choices - even compared with other forms of episodic thinking (e.g., recall of recent or remote past events; Daniel et al., 2016; Dassen et al., 2016; Stein et al., 2018; cf. Ciaramelli et al., 2019). We systematically examined these studies to test the hypothesis that a temporal shift towards the future is a contributing core component of EFT.

(v) Content Specificity. The studies on EFT and intertemporal choice differ with respect to the degree that the future imaginings were *content-specific*, that is, that they directly incorporated the future payoffs. Some studies did instruct participants to specifically imagine the impact of choosing the delayed monetary option (e.g., Benoit et al., 2011; Palombo et al., 2015) or the health-relevant consequences (Chiou & W.-H. Wu, 2017; Dassen et al., 2016). Others required their participants more generally to imagine any future event of their choosing (e.g., Cheng et al., 2012; Wu et al., 2017).

Content-specific EFT may particularly motivate farsighted decisions by explicitly binding the delayed option to one's personal future goals (Boyer, 2008; Ernst et al., 2018). However, the few studies that have directly compared the influence of content-specific versus generic EFT have yielded inconclusive results (Dassen et al., 2016; Hollis-Hansen et al., 2019; O'Donnell et al., 2017). We here made use of the methodological variance in the literature to further examine this account across all included studies.

(vi) Choice Impulsivity. An individual tendency for steeper delay discounting has been linked to maladaptive outcomes and diverse clinical disorders (Amlung et al., 2016, 2017, 2019; Paret et al., 2017). It thus is particularly important to gauge whether people who are characterized by choice impulsivity (e.g., individuals with obesity or with substance use disorder) may benefit from EFT (Peters & Büchel, 2011). Indeed, given their generally greater disregard for the future, it seems possible that their choices may in fact be more malleable to the influence of directed EFT.

Further Procedural Moderators.

In addition to the theoretically-motivated moderators, we also examined a number of additional moderators that are of more procedural interest. Specifically, we tested whether there are systematic differences in the observed effect sizes (vii) for *between- versus within-subject designs* and for studies conducted (viii) *in person* (e.g., the physical laboratory) *versus online* (e.g., MTurk). We also classified the studies to the degree that they were (ix) *prone to demand effects* (Rung & Madden, 2019) by evaluating the probability that their purpose could be discerned. Finally, in an attempt to gauge the impact of publication bias (Rothstein et al., 2005), we included a moderator analysis based on the (x) *publications status* of the experiment. This was possible due to the identification of a large number of unpublished experiments (16% of the included effect sizes).

Summary. This meta-analysis provides a comprehensive quantitative review of the impact of EFT on intertemporal choices. It examines the influence on delay discounting of monetary rewards and, for the first time, on health-relevant behavior. The targeted moderator analyses critically inform our understanding of the core components of EFT that are involved in nudging choices towards more farsighted decisions.

Method

Our meta-analysis was guided by the *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) checklist (Moher et al., 2009; Fig. 1). The full coding sheet is included in the Supplemental Material. All data and R analysis scripts are publicly available at the Open Science Framework (`https://osf.io/9rejf/?view_only=fee62b3cadd4405b8aafdac75015f60b`).

Literature Search. We sought to identify all studies that had investigated the impact of EFT on intertemporal choices involving either monetary or health-relevant decisions. We therefore searched PubMed and the Web of Science data bases for articles published before February 2020 using the following search terms:

- PubMed: (("episodic future thinking") OR ("episodic simulation") OR ("episodic prospection") OR ("mental time travel") OR ("prospective thinking") OR ("counterfactual thinking") OR (foresight)) AND (discount*)
- Web of Science: (TS=("episodic future thinking" OR "episodic prospection" OR "episodic simulation" OR "mental time travel" OR foresight) AND TS=(discount*))> OR "prospective thinking"> OR "counterfactual thinking")

In addition, we retrieved all 364 unique references cited in the identified articles on EFT and health-relevant decisions. We were further concerned that our database search had missed relevant work from a prolific lab in the field, and thus examined all of their 314 published articles. We also reviewed the reference section of recent surveys of the literature (Bulley et al., 2016; Rung & Madden, 2018), which added 7 relevant citations.

To mitigate the likely influence of publication bias, we moreover aimed to retrieve pertinent studies that had not been published in peer-reviewed journals. We first searched the data bases ProQuest and (the German-language alternative) Thesius for otherwise unpublished Master and PhD theses:

- ProQuest: (episodic AND future AND thinking NEAR discount OR episodic AND simulation NEAR discount OR episodic AND prospection NEAR discount OR mental AND time AND travel NEAR discount OR prospective AND thinking NEAR discount OR counterfactual AND thinking NEAR discount OR foresight NEAR discount)

- Thesius: (delay AND discounting AND Zukunft) as well as (delay AND discounting AND episodisch)

In addition, to identify further unpublished experiments, we emailed all 25 corresponding or senior authors of pertinent publications for which we could retrieve current contact information (on 18th of February 2020). This resulted in the additional inclusion of n = 2 studies. In total, our literature search thus yielded 1,625 hits after removing duplicates (Fig. 1).

Inclusion and exclusion criteria.
We included all studies that (i) entailed a condition that induced EFT prior to the intertemporal choices, (ii) compared this condition to a non-EFT control condition, (iii) included either monetary or health-relevant intertemporal choices, (iv) were written up in English. We did not include data of amnesic individuals, given their well-documented deficiency in EFT (McCormick et al., 2018; Race et al., 2011).

S.R. and D.F.S. independently screened the titles and abstracts. When there was any doubt about an article's eligibility, it was retained for further review. This approach led to the selection of 234 potentially relevant articles for full-text reading. Based on the above criteria, we eventually retained a total of 48 articles that contained a combined 58 studies (Fig. 1).

Data Extraction. Data were extracted and coded by S.R. and a trained research assistant (MSc level), including information required for the calculation of effect sizes (i.e., sample sizes, means and standard deviations, t-, F- or p-values). We calculated the standard deviations from their respective standard errors when only the latter were available.

Forty-three percent of the reports only presented relevant statistics as graphs. In these cases, we used the WebPlotDigitizer (Rohatgi, 2019) to extract the data to the second decimal with perfect inter-rater reliability (intraclass correlation coefficient = 1.00). For studies that did not provide sufficient information to compute effect sizes, we emailed the authors. Of all studies that had met the inclusion criteria, we thus only had to exclude a single study due to missing data.

Moderator Coding. The moderators were coded by S.R. and a trained research assistant (MSc level), who discussed and resolved discrepancies throughout the coding process. The coding criteria are detailed in Table 1.

Statistical Analysis. We computed a series of meta-analyses in R 3.6.1 (R Core Team, 2019) with the package metafor 2.4 (Viechtbauer, 2010) to assess the impact of EFT on monetary and health-relevant intertemporal choices.

Deriving Effect Sizes. We computed effect sizes as the bias-corrected Hedges' g (using the escalc function; measure option set to SMD). For *between-subjects* designs, this represents the bias-corrected standardized mean difference between the EFT and control groups. For *within-subjects* designs, we derived its equivalent (Morris & DeShon, 2002), the standardized mean change using raw score standardization (Becker, 1988; using the escalc function; measure option set to SMCR). d-values were converted via the Hedges' g function of the esc package (Lüdecke, 2019).

Determining the effect sizes for within-subject designs requires the correlation between EFT and control conditions. This correlation has rarely been reported in the surveyed literature. As recommended (Borenstein et al., 2009), based on pertinent data from our and other labs (Martínez-Loredo et al., 2017; Strickland et al., 2019; Weafer et al., 2013), we thus estimated the coefficient as r = .6. (Note that additional sensitivity analyses, using alternative correlation coefficients of r = .2 and r = .8, yielded qualitatively identical results. See Supplemental Material).

Some outcome measures were reverse-coded, i.e., greater numbers reflected less patient choices (e.g., k-values as in Mellis et al., 2019). In these cases, we multiplied the reported value by -1 to ensure that positive effect sizes consistently represented more farsighted decisions. All effect sizes are reported in Table 2.

Data Synthesis. Seventy-seven percent of the reports reported multiple effect sizes, including multiple studies per report (applicable for 17% of the reports), outcomes for both monetary and health-relevant choices, different dependent variables for the same outcome (e.g., k and percentage of immediate choices), and different groups (e.g., two EFT groups versus the same control group).

The data thus entail two types of dependencies: *correlated effects* (i.e., effect sizes derived from overlapping samples) and *hierarchical effects* (i.e., effect sizes nested within studies; Tanner-Smith & Tipton, 2014). These dependencies violate assumptions underlying traditional meta-analytic approaches (Lipsey & Wilson, 2001) and could artificially narrow confidence intervals and shrink standard errors, and, as a consequence, increase type I error (Hox, 2010; Viechtbauer, 2010).

We employed two methods to account for both types of dependencies while retaining all provided information and without losing statistical power (see Cheung, 2014, 2019). First, due to the hierarchical dependency, we adopted a three-level random-effects meta-analytic model. This model accounts for the variance in the observed effect sizes (level 1),

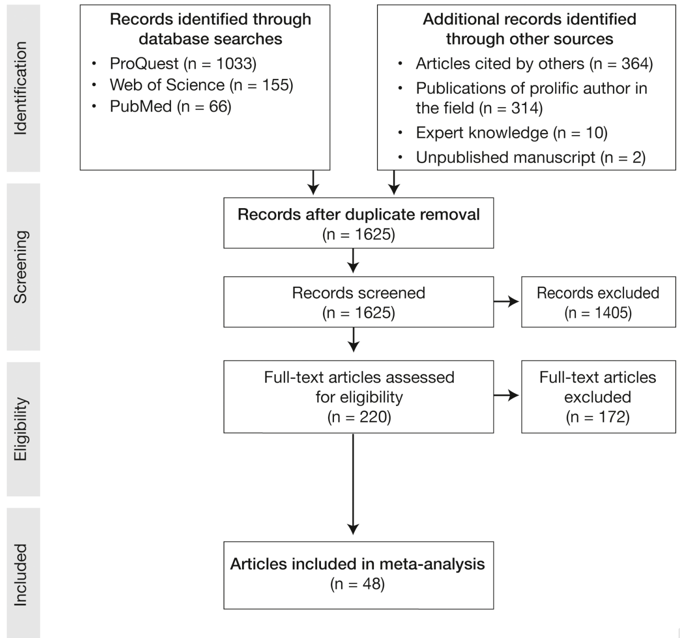

Fig. 1. *PRISMA Flowchart Illustrating Study Selection.*

variance between effect sizes within a report (level 2), and variance between reports (level 3; Cheung, 2014). This approach allowed us to estimate the random variation at level 2 to account for dependence (i.e., clustering of effect sizes) among effect sizes from the same article (Konstantopoulos, 2011). Model parameters in the three-level model were estimated using restricted maximum likelihood estimation (REML; Cheung, 2014; Viechtbauer, 2010) with the Knapp & Hartung (2003) method for calculating regression coefficients and confidence intervals (Assink & Wibbelink, 2016).

Second, due to the correlated dependencies, we performed statistical tests on cluster-robust standard errors and confidence intervals (Hedges et al., 2010) that we generated for the estimates from the three-level meta-analytic model (via the robust function of the metafor package; Viechtbauer, 2010). Individual coefficients were tested using a t-distribution and all model coefficients were tested using an F-distribution (i.e., omnibus test; excluding the intercept).

Assessment of Heterogeneity. We evaluate the heterogeneity of the reported models by reporting total I^2-values along with the components for the second (I^2_{Level2}; variance within the same studies) and third level (I^2_{Level3}; variance between studies). These values indicate the percentage of variability that is based on differences between studies rather than sampling error. Heterogeneity thus provides information regarding the influence of study-level characteristics (e.g., valence of the imagination) on the observed effect size. We classified heterogeneity, following the Cochrane Handbook (Higgins & Green, 2011), with I^2-values above 50% as substantial and above 75% as considerable.

We determined the significance of the heterogeneity at each level using the Q-statistic. This approach entailed separate one-tailed log-likelihood-ratio tests in which the deviance of the full model was compared to the deviance of the model excluding one of the variance parameters (Assink & Wibbelink, 2016). A significant Q-value at either of the levels suggests systematic differences beyond what would be expected from sampling error alone, potentially due to significant moderating factors accounting for heterogeneity.

Moderator Analysis. An important aim of our analysis was the assessment of the critical core components that influence the impact of EFT on intertemporal choices. We thus took advantage of the heterogeneity of the reported effect sizes and conducted a series of moderator analyses. These were targeted at hypothesized candidate components and at further procedural factors. The potential moderators varied either between and within studies (e.g., monetary versus health-relevant choices) or between studies (e.g., publication status).

We dummy-coded all categorical moderators with more than two levels (e.g., for the moderator valence) before entering them into meta-regression equations. Due to the limited number of studies that contributed to the levels of the moderators, we focused on a series of single-moderator analyses. We also report a model that included all moderators with significant univariate effects to reduce multicollinearity (Hox, 2010). Moderator analyses need a large amount of observations and low heterogeneity to achieve high power (Hedges & Pigott, 2004). Null effects in tests of moderation should thus be interpreted cautiously.

Model Comparisons. We informed our moderator analyses by assessing whether each model provided a better fit to the data than a more parsimonious intercept-only model. These model comparisons were based on the Akaike Information Criterion (AIC; Akaike, 1973) with small-sample correction.

The AIC values were transformed to conditional probabilities for each model, i.e., AIC weights (AICw; Wagenmakers & Farrell, 2004). This was done using the AIC (from the metafor package) and the Weights (from the MuMIn package) functions in R.

However, AIC is not suitable for comparing models that differ in their fixed effects structure and that have been estimated via REML. Therefore, we based these comparisons on models that were refitted using the Maximum Likelihood method, again with the Knapp & Hartung (2003) correction.

To obtain a measure of statistical significance, we further compared the single-moderator models using Likelihood Ratio Tests. These outperform omnibus moderator tests in terms of statistical accuracy (i.e., Type I error rates; López-López et al., 2017).

Assessment of Publication Bias. Meta-analyses are prone to provide exaggerated effect size estimates due to publication bias, i.e., the selective reporting of studies yielding significant effects (Carter et al., 2019; Renkewitz & Keiner, 2019). To minimize this bias, we sought to identify studies without a peer-reviewed publication record. Notably, 16% of our effect sizes stem from such unpublished sources. In addition, we (i) examined the degree to which our results remain bias-inflated and (ii) attempted to adjust for such effects. To accomplish these goals, we employed several complementary methods, given that none of them is without their particular drawbacks (Carter et al., 2019; Renkewitz & Keiner, 2019) or capable to remedy all questionable research practices (Simonsohn et al., 2014).

The first set of analyses was directly based on the three-level model: *First*, we used publication status as a moderator to examine whether there is a systematic difference in the reported effect sizes. *Second*, we visually inspected contour-enhanced funnel plots for asymmetries that are likely indicative of publication bias: lack of non-significant findings to the left side of the meta-analytically estimated effect size and an overrepresentation of low-powered findings to its right side. Visual inspection, however, is inherently subjective in its interpretation. Moreover, because the plots do not consider the data's three-level-structure, clusters of data points and ensuing asymmetry may be mistaken as bias (Lau et al., 2006).

Third, to formally assess funnel plot asymmetry, we conducted Egger's test (with $p < 0.1$ as the critical value; Egger et al., 1997). This test evaluates whether the precision of the effect (i.e., the standard error) is associated with the size of the effect. We adapted it to the three-level meta-analytic model by adding random effects for between- and within-study variance to account for dependency and by including the precision of the effect as the informative moderator (Fernández-Castilla et al., 2019). A statistically significant slope indicates bias by suggesting that effect sizes are systematically different for high- versus low precision studies. *Fourth*, we similarly interrogated the funnel plot using the Macaskill regression test that examines whether effect sizes can be regressed on sample sizes (Macaskill et al., 2001). Though this test has fairly low power, it typically outperforms Egger's test in controlling Type I error (Kromrey & Rendina-Gobioff, 2006; Macaskill et al., 2001).

A second set of analyses required that the included effect sizes are all independent (Assink & Wibbelink, 2016). Thus, they could not be based on our three-level model. For the following analyses, we ensure independence by aggregating all dependent effect sizes (via the agg function of the MAd package; del).

The first two methods examine the distribution of p-values. They are based on the rationale that the distribution of significant p-values contains more low- ($p \leq .01$) than high- ($p > .04$) p-values for a true effect that is not influenced by publication bias.

First, we applied p-curve (Simonsohn et al., 2014) via the pcurve function of the dmetar package (Harrer et al., 2019). We estimated the true effect (effect.estimation option set to TRUE) with the search space set between $g = 0.00$ and $g = 1.0$. We report results for the "full" (for all p-values $< .05$) and "half" p-curve (for all p-values $< .025$). The half p-curve is more sensitive to stronger cases of p-hacking, though at the cost of reduced statistical power (Simonsohn et al., 2015). P-curve attempts to correct for publication bias by computing the probabilities of observing the significant reported effect sizes contingent on the meta-analytical effect size being the true underlying one. Importantly, in the case of highly heterogeneous data, the estimates provided by p-curve should be considered estimates of the average true effect sizes of only the statistically significant set of studies (van Aert et al., 2016).

Second, we implemented p-uniform* (van Aert & van Assen, 2020) via the puni_star function of the puniform package (van Aert, 2019). This method, in contrast to p-curve, also takes nonsignificant effects into account and thus allows for a direct estimation and test of (full) between-study variance. It thus attempts to safeguard against the overestimation of the effect size that could arise from including significant effects only and, more generally, as a consequence of between-study variance (Carter et al., 2019; van Aert et al., 2016).

Finally, we created a funnel plot of the aggregated effect sizes and applied the trim-and-fill method (Duval & Tweedie, 2000). Trim-and-fill seeks to restores the symmetry of the funnel plot by imputing presumably missing effect size estimates. These are then entered into a new random-effects meta-analytical model. We note, however, the current debate regarding the validity of this method (Carter et al., 2019; Renkewitz & Keiner, 2019; van Aert et al., 2016), and thus suggest to interpret the results with caution (Sutton et al., 2000).

Results

Descriptive Statistics. The current meta-analysis included 48 articles that reported 58 independent studies and comprised a total of 174 effect sizes. They were published between 2010 and 2020. The overall sample size was N = 3,882, with individual sample sizes ranging from n = 8 (Cooper, 2013) to n = 200 (Bulley et al., 2019). Most articles examined monetary intertemporal choices only (n = 32, 67% of effect sizes), some included both monetary and health-relevant choices (n = 13, 27%), and a few were focused on health-relevant choices only (n = 3, 6%).

Overall Effect. Overall, we obtained a significant, medium-sized effect, g = 0.439, SE = 0.053, t = 8.324, $p < .001$, 95% confidence interval (CI) = [0.333, 0.545], with substantial overall heterogeneity, $Q_E(173) = 684.42$, $p < .001$, I^2 = 79.21, 95% CI [72.48, 85.23] (Fig. 2).

Variance of the Overall Effect Size. We next sought to examine the distribution of the variance across the three-levels of our meta-analytical model. The choice of this model was corroborated by an intraclass correlation coefficient of ρ = .483, indicating that the underlying true effects within reports were strongly correlated. Moreover, the variance was significant at both the within-study level (estimate = .083, $p < .001$) and the between-study level (estimate = .077, $p < .001$).

Follow-up analyses indicated that variance was distributed across the sampling, within-study, and between-study levels as 20.79%; 40.94%, 95% CI [26.47, 61.89]; and 38.27%, 95% CI [16.34%, 78.47%], respectively. The low percentage of variance attributed to the samples further motivates the exploration of potential moderators (Hunter & Schmidt, 1990).

Moderator Analyses. An important goal of the present meta-analysis was the examination of possible core components of EFT that influence intertemporal choices. These moderator analyses were clearly further motivated by the substantial heterogeneity and low contribution of *samples* to the effect-size variance. Specifically, we report a series of moderator analyses that examined theoretically-motivated candidate core components (Fig. 3) as well as additional, more procedural factors of the employed study designs (Fig. 4). First, however, we examine whether EFT is differentially effective in influencing monetary versus health-relevant intertemporal choices. (All moderator results are further detailed in Table 3.)

Choice Domain. The model for the moderator *choice domain* (monetary versus health-relevant), AICw = 33.3%, $\chi^2(2)$ = 0.71, p = .401, did not show a better fit to the data than the intercept-only model. Accordingly, the effect of choice domain was not significant, $F(1, 46) = 0.98$, p = .328, $Q_E(172)$ = 681.75 $p < .001$. This analysis thus provides no evidence that EFT has a stronger impact on either kind of intertemporal choice. Indeed, consistent with our hypothesis, we obtained significant effect sizes for both monetary, g = 0.462, $t(46)$ = 7.83, $p < .001$, and health-relevant choices, g = 0.385, $t(46)$ = 5.05, $p < .001$.

Theoretically-Motivated Moderators.

(i) Valence. We had hypothesized that EFT would have the strongest impact on intertemporal choices when the simulated episodes are positive, and possibly the weakest – or even reversed – impact when they are negative in nature. Indeed, the model including the valence moderator showed a superior fit to the data, carrying approximating 100% of the cumulative model weight. It was thus significantly better than the intercept-only model, $\chi^2(3)$ = 29.21, $p < .001$.

Consistent with these results, the main effect of valence was significant, $F(3, 44) = 3.02$, p = .04, $Q_E(170)$ = 555.73, $p < .001$. As predicted, positive episodes yielded the largest and significant effect, g = 0.52, $t(44)$ = 8.79, $p < .001$, followed by positive-to-neutral episodes, g = 0.438, $t(44)$ = 4.37, $p < .001$, with only a trend for neutral episodes, g = 0.184, $t(44)$ = 1.83, p = .074. Indeed, negative episodes were associated with the smallest, and non-significant effect, g = -0.179, $t(44)$ = -0.52, p = .607.

Our predictions were further largely corroborated by the comparisons between the levels of the moderator. The effect for positive episodes was significantly larger than for neutral episodes, β = 0.336, $t(44)$ = 2.96, p = .005, and marginally larger than for negative episodes, β = 0.699, $t(44)$ = 1.94, p = .059. Positive-to-neutral events yielded trends for a larger effect compared with neutral, β = 0.254, $t(44)$ = 1.8, p = .08, and negative episodes, β = 0.617, $t(44)$ = 1.71, p = .094. For all other direct comparisons, t was smaller than 1.29, p greater than .2 (Fig. 3A).

(ii) Vividness. We next tested the hypothesis that imagined events should be particularly impactful when they are experienced as more vivid. We thus determined the difference in vividness between the EFT and respective control condition and used the effect sizes as a moderator. (Note that we could not compare this model with the intercept-only model, because it was only based on the subsample of studies that had provided vividness ratings.)

As predicted, the moderator analysis yielded the significant main effect of vividness, $F(1, 13) = 8.43$, p = .012, $Q_E(58)$ = 126.51, $p < .001$. Across studies, a vividness increase of one standard deviation was associated with an increase of β = 0.180, $t(13)$ = 2.91, p = .012, in the standardized mean difference (Fig. 3B).

(iii) Episodicity. We also examined whether the episodic nature of the future simulations constitutes a core component of EFT with respect to its influence on intertemporal choices. We thus performed a moderator analysis with two groups: One group included effect sizes that compared an EFT condition with control conditions that were closely matched in

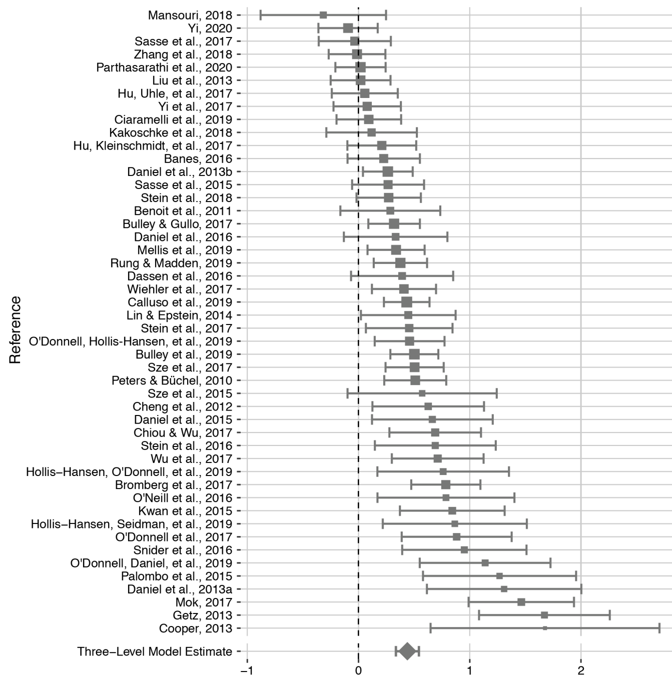

Fig. 2. *Forest plot on Aggregated Effect Sizes.* Symbols indicate the respective effect size (Hedges' g). Larger symbols indicate smaller sampling variance (i.e., larger sample sizes). Bars denote 95% Confidence Intervals.

all aspects but the episodicity (i.e., studies requiring semantic future thinking). The other group included the effect sizes for comparisons with all other control conditions. The corresponding moderator model did not show a significantly better fit to the data than the intercept-only model, AICw = 29.5%, $\chi^2(1) = 0.35$, $p = .552$, and there was accordingly no main effect of the moderator, $F(1, 46) = 0.77$, $p = .384$, $Q_E(172) = 678.38$, $p < .001$. Critically, however, we corroborated that we did not only obtain a significant effect for the more disparate control conditions, $g = 0.433$, $t(44) = 8.10$, $p < .001$, but also for those that only differed from EFT in their lack of episodicity, $g = 0.553$, $t(44) = 4.15$, $p < .001$, $\beta = 0.12$, $t(45) = 0.88$, $p = .383$ (Fig. 3C).

(iv) Future-Orientedness. Analogously to the previous analysis, we sought to determine the contribution of the temporal orientation. We therefore split the effect sizes based on whether the control condition was closely matched in all aspects (including the episodicity) but the future-orientedness (e.g., episodic recent thinking) or whether the control conditions were more disparate. This model did not fit the data better than the intercept-only model, AICw = 26.7%, $\chi^2(1) = 0.08$, $p = .779$ and there was no main effect of the moderator, $F(1, 46) = 0.07$, $p = .789$, $Q_E(172) = 663.1$, $p < .001$. Importantly though, we observed not only a significant effect for the more varied non-episodic control conditions, $g = 0.45$, $t(46) = 6.51$, $p < .001$, but also for those that were matched in all facets but the future-orientedness of the imagination, $g = 0.430$, $t(46) = 6.57$, $p = <.001$, $\beta = 0.\text{-}0.02$, $t(45) = 0.27$, $p = .789$ (Fig. 3D).

(v) Content Specificity. We next tested the hypothesis that EFT exerts a stronger influence on intertemporal choices if the simulation directly incorporates the future pay-off (e.g., the enjoyment of the later reward). Indeed, the model for *content specificity* (content-specific versus general EFT) showed a good fit to the data, carrying 90.5% of the cumulative model weight, and outperformed the intercept-only model, $\chi^2(1) = 6.61$, $p = .01$. The moderator was significant, $F(1, 46) = 9.21$, $p = .004$, $Q_E(172) = 657.19$, $p < .001$, reflecting that the effect was stronger for content-specific, $g = 0.707$, $t(46) = 6.95$, $p < .001$, $\beta = 0.317$, $t(45) = 3.03$, $p = .004$ than general future simulations, $g = 0.391$, $t(44) = 7.46$, $p < .001$ (Fig. 3E).

(vi) Choice Impulsivity. We finally tested the hypothesis that the impact of EFT would be larger for individuals characterized by choice impulsivity than for neurotypical individuals. If the former tend to put less weight on the future consequences of their actions, they may particularly benefit from such an intervention. The moderator model did fit the data significantly better than the intercept-only model, AICw = 76.8%, $\chi^2(1) = 4.49$, $p = .034$, and choice impulsivity had a significant main effect on the effect size, $F(1, 46) = 6.23$, $p = .016$, $Q_E(172) = 639.37$, $p < .001$. Even though neurotypical samples do benefit from EFT, $g = 0.367$, $t(44) = 6.44$, $p < .001$, the effect was significantly larger for individuals characterized by choice impulsivity, $g = 0.604$, $t(45) = 8.02$, $p < .001$, $\beta_1 = 0.237$, $t(45) = 2.5$, $p = .016$ (Fig. 3F).

However, we observed the effect of choice impulsivity across monetary and health-relevant choices (e.g., alcohol consumption). The latter choices were clearly more relevant for the impulsive groups (e.g., people with alcohol substance abuse) than for their respective controls. We thus wanted to establish whether this moderator influences both, pertinent health-relevant and also general monetary choices.

For health-relevant choices only, *choice impulsivity* was

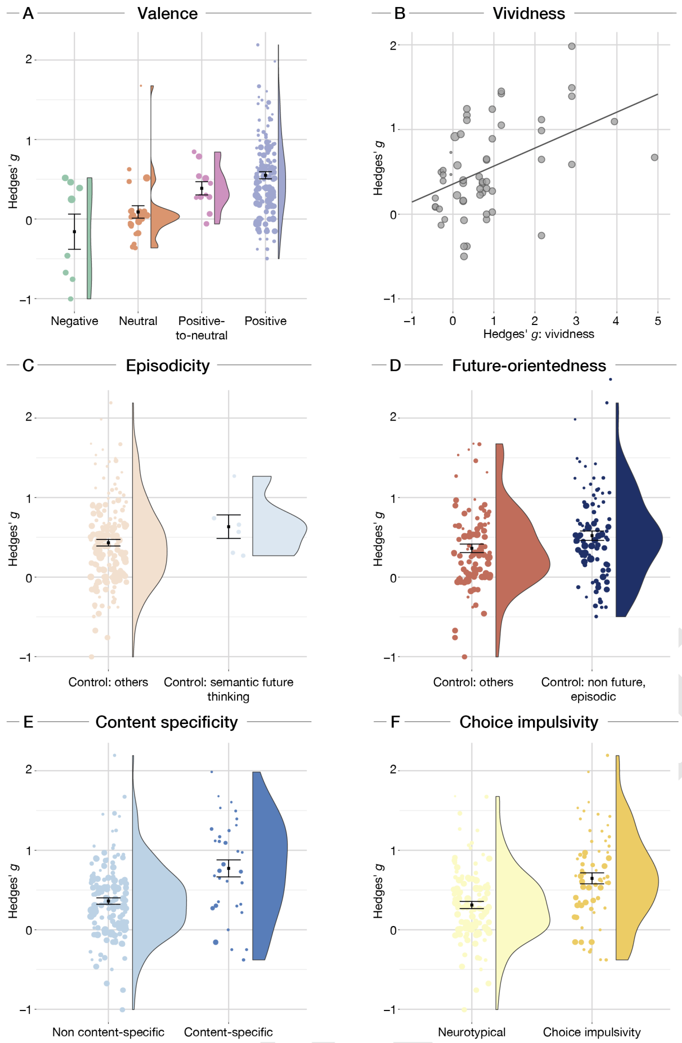

Fig. 3. *Effect Sizes (Hedges' g) of the Theoretically-Motivated Moderators.* Each dot shows the observed effect size for a single study for the respective moderator level. Studies with larger samples (i.e., greater precision) are displayed with larger dots. Black squares indicate the mean, whiskers the standard error. The shape illustrates the distribution of effect sizes for the respective moderator level. Panel A: Valence. Panel B: Vividness. Panel C: Episodicity. Panel D: Future-orientedness. Panel E: Content specificity. Panel F: Choice impulsivity.

again a significant moderator, $F(1, 14) = 12.58$, $p = .003$, $Q_E(69) = 189.43$, $p < .001$. Samples characterized by choice impulsivity did make more health-beneficial choices following EFT, $g = 0.572$, $t(14) = 6.64$, $p < .001$, $\beta_1 = 0.445$, $t(14) = 3.55$, $p = .003$, whereas the neurotypical samples did not show such an effect, $g = 0.128$, $t(14) = 1.24$, $p = .22$.

For monetary choices only, *choice impulsivity* constituted just a marginal moderator, $F(1, 43) = 3.67$, $p = .062$, $Q_E(101) = 424.17$, $p < .001$. This reflected a trend for a greater effect in samples characterized by choice impulsivity, $g = 0.619$, $t(45) = 6.8$, $p < .001$, $\beta_1 = 0.21$, $t(42) = 1.91$, $p = .062$. Importantly, for the monetary choices, the neurotypical individuals also showed a significant effect, $g = 0.41$, $t(42) = 6.32$, $p < .001$.

Procedural Moderators. In the following, we test possible moderators that reflect major procedural differences rather than theoretically-relevant components of EFT.

(vii) Design. We tested whether effect sizes would differ for studies that had employed between- versus within-subject manipulations. The model for design, AICw = 72.2%, $\chi^2(1) = 4.0$, $p = .045$ showed a significantly better fit to the data than the intercept-only model, though the moderator was only marginally significant, $F(1, 46) = 3.8$, $p = .057$, $Q_E(172) = 621.71$, $p < .001$. There was a trend for effect sizes from between-subjects designs, $g = 0.509$, $t(46) = 7.26$, $p < .001$ to exceed those from within-subjects designs, $g = 0.319$, $t(46) = 4.89$, $p < .001$, $\beta_1 = 0.19$, $t(45) = 1.95$, $p = .057$ (Fig. 4A).

(viii) Study Site. The model comparing experiments conducted in person versus online did not fit the data better than the intercept-only model, AICw = 28.4%, $\chi^2(1) = 0.24$, $p = .623$, and was not a significant moderator, $F(1, 46) = 1.02$, $p = .278$, $Q_E(172) = 678.74$, $p < .001$. Both in-person experiments, $g = 0.451$, $t(46) = 7.50$, $p < .001$, and experiments conducted online, $g = 0.377$, $t(46) = 8.27$, $p < .001$ yielded significant effects (Fig. 4B).

(ix) Proneness to Demand Effects. The model that in-

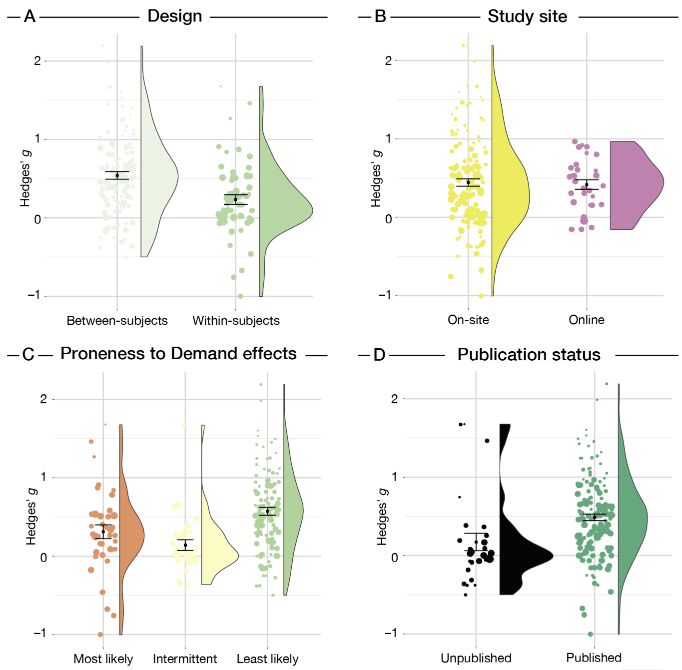

Fig. 4. Effect Sizes (Hedges' *g*) of the Procedural Moderators. Each dot shows the observed effect size for a single study for the respective moderator level. Studies with larger samples (i.e., greater precision) are displayed with larger dots. Black squares indicate the mean, whiskers the standard error. The shape illustrates the distribution of effect sizes for the respective moderator level. Panel A: Design. Panel B: Study site. Panel C. Proneness to Demand effects. Panel D: Publication status.

cluded the proneness of studies for demand effects exhibited a significantly better fit to the data than the intercept-only model, AICw = 77.3%, $\chi^2(1) = 6.67$, $p = .036$. The moderator was also significant, $F(2, 45) = 3.97$, $p = .026$, $Q_E(171) = 597.01$, $p < .001$ (Fig. 4C). Critically, we observed that those studies that were the *least likely* to be affected by demand effects actually showed the numerically largest effect size, $g = 0.532$, $t(45) = 8.5$, $p < .001$. This effect size was larger than the one for studies with an *intermittent* chance of inferring the experimenter's intent, $g = 0.201$, $t(45) = 1.97$, $p = .055$, $\beta = 0.331$, $t(45) = 2.76$, $p = .008$, but did not differ from effect sizes where the study intent was *most likely* to be inferred, $g = 0.381$, $t(45) = 3.97$, $p < .001$, $\beta = 0.151$, $t(45) = 1.32$, $p = .194$. There was no significant difference in effect size between the latter type of studies and the intermittent ones, $\beta = 0.18$, $t(45) = 1.28$, $p = .207$. Thought this moderator was significant, it thus indeed yielded evidence *against* demand effects.

(x) Publication Status. The model that differentiated between published and unpublished studies did not fit the data better than the intercept-only model, AICw = 38.5%, $\chi^2(1) = 1.16$, $p = .282$ – consistently, publication status was also not a significant moderator, $F(1, 46) = 0.55$, $p = .463$, $Q_E(172) = 639.16$, $p < .001$ (Fig. 4D). These data thus do not provide evidence for a selective reporting of results based on the magnitude of the effect size. However, on their own, only published, $g = 0.457$, $t(45) = 9.37$, $p < .001$, but not unpublished studies, $g = 0.307$, $t(45) = 1.53$, $p = .132$, did show a significant effect.

Multiple Moderator Model. We finally fit a multiple moderator model that included all significant moderators (except for the vividness moderator that was fit on only a subset of the data). As summarized in Table 4, the omnibus test was significant, $F(7, 40) = 3.68$, $p = .004$, $Q_E(166) = 500.22$, $p < .001$, suggesting that at least one of the regression coefficients of the moderators deviated significantly from zero. Accounting for multicollinearity (Hox, 2010), these findings revealed that content-specific (versus general) EFT had a significant unique moderating effect on the association between EFT and decision-making, $\beta = 0.245$, $t(45) = 2.17$, $p = .036$. In addition, there were trends for positive (versus neutral), $\beta = 0.265$, $t(45) = 2.02$, $p = .051$, and positive-to-neutral (versus neutral) episodes, $\beta = 0.25$, $t(45) = 1.89$, $p = .066$, to have a greater positive impact.

Assessment of Publication Bias. We sought to minimize the impact of publication bias on our meta-analytical results by including studies that had not been published in peer-reviewed journals. The analysis provided no evidence for a moderating effect of publication status. However, we used various complementary methods to further assess potential publication bias. (See Supplemental Material for bias corrections separately for monetary and health-related choices.)

Assessment on the Three-Level Model. The first set of analyses was based on a funnel plot of our main analytical approach, i.e., the three-level model. Egger's regression test provided evidence for significant asymmetry, $F(1, 46) = 27.64$, $p < .001$, indicating publication bias. By contrast, this was not the case for the Macaskill regression test, $F(1, 46) = 0.64$, $p = .43$.

Assessment on Aggregated Effect Sizes. The second set of analyses was based on aggregated effect sizes. Visual inspection of the funnel plot suggested publication bias (Fig. 5A. We followed-up on this impression using various formal statistical tests:

p-curve. If there is evidential value in the pattern of p-values, the corresponding p-curve should be right-skewed (i.e., towards values $\leq$ .01; Fig. 5 B). Indeed, both the tests for the half (i.e., $p < .025$; $p_{Half} < 0.001$) and full p-curve (i.e., $p < .05$, $p_{Full} < 0.001$) were significant. Moreover, the p-curve was also not flat (tests for flatness: $p_{Full} = 1.0$, $p_{Binomial}$= 1.0). The power estimate of the examined studies was high (87% [95% CI= 76.7% - 93.1%]) and the true effect size was estimated as $g = 0.677$. However, given the significant heterogeneity (> 50%), this estimate may not be considered trustworthy (Harrer et al., 2019; van Aert et al., 2016).

p-uniform*. P-uniform* indicated significant between-study variance, τ^2 =.076 [.035, .150], $L = 35.42$, $p < 0.001$, though the publication bias test was not significant, $L = 2.42$, p= .298. This approach thus does not provide evidence for the presence of publication bias. p-uniform* estimated the effect size as $g = 0.427$ [.268, .579], $L = 16.25$, $p < .001$.

Trim-and-fill. The trim-and-fill procedure (Duval & Tweedie, 2000) evens out the symmetry of the funnel plot by imputing additional values and trimming the values of extreme observations. It estimated that 13 studies were missing to achieve symmetry (Fig. 5A). However, following adjustment for this putative bias, the newly estimated meta-analytic effect size remained significant at $g = 0.299$, 95% CI [.166, .431].

Summary. The methods did not all converge with respect to the question whether our results are still affected by publication bias following the inclusion of the identified unpublished results. Egger's test and trim-and-fill did indicate publication bias, whereas this was not the case ffor the moderator analysis, the Macaskill regression test, as well as p-curve and p-uniform*. Critically, however, the impact of EFT on intertemporal choices remained significant following all attempts to adjust for publication bias.

Discussion

Our pervasive tendency to devalue future rewards often leads to myopic decisions. We forego larger amounts of money in favor of smaller short term gains or seek the instant pleasures of rich desserts at the cost of our long-term well-being. We here performed the first comprehensive and quantitative analysis of the hypothesis that episodic future thinking can positively influence such intertemporal choices. Specifically, in a three-level model, we meta-analyzed the rich literature that has accumulated over the last decade with a total of 174 effect sizes. Our results indicate that EFT has a statistically significant, medium-sized effect on both monetary and health-relevant intertemporal choices. Notably, 16% of our effect sizes stemmed from unpublished articles and the effect stayed stable following further attempts to adjust for remaining publication bias. Overall, this meta-analysis thus corroborates that EFT promotes farsighted decisions.

The meta-analytical model also yielded significant and large heterogeneity in the effect sizes, hinting at meaningful experimental differences that do not reflect simple sampling error. We exploited the heterogeneity in a number of moderator analyses. These assessed critical core components of EFT that may be instrumental in influencing intertemporal choices as well as more procedural study differences.

Core Components of Episodic Future Thoughts.

Valence. The impact of EFT varied considerably with the valence of the prospective episode. As hypothesized, we obtained a graded pattern, with the largest beneficial effect for positive events, followed by positive-to-neutral events. Neutral episodes only yielded a trend and negative episodes were also not significant.

On the one hand, EFT acts as an "affective forecast" (Gilbert & Wilson, 2007) that conveys the anticipated positive affect that the delayed reward would hold. This *anticipated affect* may add to the valuation of the delayed reward and thus effectively attenuate its discounting (Benoit et al., 2018; Berns et al., 2007; Boyer, 2008; Bulley et al., 2016; Bulley & Schacter, 2020). EFT can thus make the reward feel worth the wait. On the other hand, EFT further conveys utility to the wait itself by also eliciting *anticipatory affect*. For example, there is evidence that people opt to defer events that they deem particularly pleasurable, such as a promised kiss from a movie star – presumably because they also cherish the prospect in and of itself (Loewenstein, 1987).

EFT thus encourages individuals to organize their behavior in accordance with anticipated and anticipatory positive emotions (Baumgartner et al., 2008; Lempert et al., 2016). However, given that the few studies instructing for negative EFT did not yield a significant effect, the data did not support the account that imagining a negative future tethers individuals to the immediate present (Liu et al., 2013).

Vividness. The experiments in which participants imagined the events more vividly also yielded the greatest impact of EFT on intertemporal choices. However, this analysis was based on only 31% of the articles (34% of effect sizes) that had reported vividness ratings for the EFT and control conditions. Though inferences from across-sample correlations are susceptible to ecological fallacy, this finding is consistent with a few studies that had reported such associations across subjects (Ciaramelli et al., 2019; Peters & Büchel, 2010).

There are a number of ways in which vividness could influence the impact of EFT. First, all of the included studies required participants to imagine positive future episodes. A greater vividness may thus have boosted the effect by eliciting a stronger positive affect (D'Argembeau & Van der Lin-

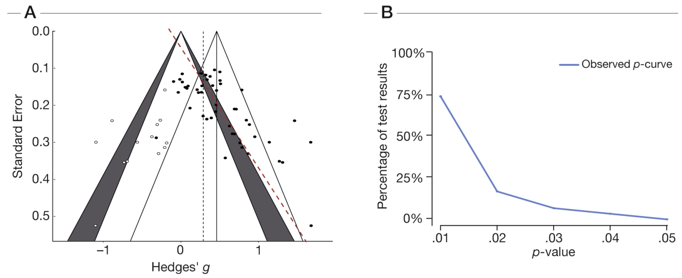

Fig. 5. *Assessing Publication Bias on Aggregated Effect Sizes.* Panel A: Funnel plot of effect size estimates and standard errors for aggregated effect sizes. Solid black dots represent effect sizes included in the analysis and empty white dots represent studies imputed via trim-and-fill (Duval & Tweedie, 2000). The dashed line shows the trim-and-fill adjusted effect size. The left funnel is centered at 0, which is the value under the null hypothesis of no effect, whereas the right funnel is centered at the model estimate under the null hypothesis that the meta-analytic model estimate is true. Contours represent 95% confidence intervals. The dashed red line shows the regression line from *Egger's test*. Panel B: Observed p-curve of all articles, including 30 statistically significant (p < .05) results, of which 29 had a p-value < .025. An additional 18 results were excluded from the p-curve because they had p-values > .05.

den, 2012; Holmes & Mathews, 2010; Pearson et al., 2015; Schubert et al., 2020). Second, imagining an event with greater vividness makes it seem more certain to occur (Gregory et al., 1982; Sherman et al., 1985; Szpunar & Schacter, 2013; J. Q. Wu et al., 2015). As a consequence, the delayed reward would seem more likely to manifest and accordingly constitute a safer bet (Bulley et al., 2016). Finally, more vivid EFT imbues the future option with more concrete episodic details, and thereby leads to a construal at a level that typically characterizes greater immediacy (D'Argembeau & van der Linden, 2004; Lempert & Phelps, 2016; Trope & Liberman, 2010). EFT may thus effectively make the future appear closer in time.

Content Specificity. The analysis revealed that the content of the future imagining is a key feature determining the impact of EFT on DD. This moderator provided a unique effect that was not accounted for by either the valence or vividness of the simulation. Specifically, the influence of EFT was greater when the imagined content was closely tied to the nature of the delayed choice option. This could either entail the future moment of spending the monetary reward (e.g., Benoit et al., 2011; Palombo et al., 2015) or the direct consequences of avoiding behavior that would be detrimental to one's health (e.g., positive life events after successfully quitting smoking; Chiou & W.-H. Wu, 2017).

The content specificity of EFT seems particularly important for the latter kind of choices, where the adverse long-term consequences of impulsive decisions can seem fairly intangible (e.g., protracted health issues). More generally, EFT has been argued to foster farsighted decisions by supporting a model-based choice process that involves the mental construction of the context in which the delayed reward would be delivered (Kurth-Nelson et al., 2012). This process thus requires a correspondence between the specific content of the imagination and the nature of the delayed choice option.

Episodicity. The analysis also yielded a significant effect for the few studies whose control conditions were tightly matched in all aspects but the imaginings' episodicity. The episodic nature of EFT thus contributes to its effect beyond what is provided by semantic future thinking (Atance & O'Neill, 2001; Szpunar et al., 2014). This finding is consistent with reports that amnesic patients do not additionally profit from cues prompting them to engage in EFT (Kwan et al., 2015; Lebreton et al., 2013; Palombo et al., 2015) even though they can attenuate DD via a more semantic mechanism that is also directed towards the future (Palombo et al., 2016). EFT thus does not merely influence intertemporal choices by inducing a generic future orientation (Lin & Epstein, 2014; Rung & Madden, 2018). Instead, the data indicate that the autonoetic experience of a future episode constitutes a critical component (Benoit et al., 2018).

Future-Orientedness. Complementary to the previous section, we also observed that EFT nudges towards farsighted choices when the control condition is matched on all aspects but the temporal direction. That is, EFT yielded an effect even when compared with other episodic conditions that were directed towards the distant (Banes, 2016; Ciaramelli et al., 2019) or recent past (Daniel et al., 2013; Rung & Madden, 2019) or towards the present (Yi, 2020). This finding dovetails with the observed effect on content specificity, in the sense that an orientation towards the future also aligns the imagined episode closer with the potential moment of receiving the later reward. The temporal direction may thus constitute a core component of EFT that shifts attention towards a delayed choice option (Lin & Epstein, 2014) and thereby provides motivational incentives for the pursuit of long-term goals (Dreves & Blackhart, 2019).

Choice impulsivity. Finally, we also established that the effect of EFT was particularly strong in groups characterized by choice impulsivity (Hamilton et al., 2015). These included individuals with obesity and substance dependences (e.g., nicotine addiction). The effect in these groups surpassed the one in neurotypical individuals. This pattern is consistent with a previous observation, albeit in a small sample, that EFT yields a greater benefit for people who are usually less prone to consider the future consequences of their actions (Benoit et al., 2011).

The effect of choice impulsivity was particularly pronounced for health-relevant decisions. This may simply reflect the more limited relevance of these decisions for the respective neurotypical control samples. For example, reducing calories intake would clearly have a greater importance for individuals with obesity than for people with normal weight. In fact, neurotypicals make, by definition, less impulsive choices, leaving less room for EFT to nudge them towards more farsighted decisions. It is thus particularly noteworthy that we observed a similar trend for monetary choices that are of comparable relevance for either group. Our results thus indicate that EFT constitutes a candidate mechanism for altering maladaptive, impulsive behavior (Levens et al., 2019; Peters & Büchel, 2011; Rung & Madden, 2018).

Summary. The foregoing analyses yielded a number of core components that contribute to the impact of EFT on discounting: valence, vividness, episodicity, future orientedness, and content specificity. The results motivate future research targeted at further delineating the interactions and unique influences of these components, possibly with a particular emphasis on optimizing intertemporal choices in individuals characterized by maladaptive choice impulsivity.

Procedural Moderators. We further conducted a number of moderator analyses targeted at salient study differences. There was no difference in effect sizes for studies that had been conducted *in person* (e.g., the laboratory) versus *online* (e.g. via MTurk). This result is promising, given the potential of online experiments to include more diverse and thus more representative samples (Paolacci & Chandler, 2014). However, we caution that not all core components may be equally amenable to online experiments. For example, given the typically shorter time frame, it seems more difficult to implement *within-subject* designs with subtle manipulations such as episodic versus semantic future thinking.

Notably, the model including *between- versus within-subject design* could account better for the data than the intercept model. *Between-subject* designs trended to yield greater effect sizes than *within-subject* designs. This is noteworthy, given that the former generally tend to be less powerful (Lakens, 2013). That finding highlights the difficulties inherent to within-subject designs (Greenwald, 1976) where the effect of EFT may carry over into the control condition. This is particularly the case when both conditions repeat the identical intertemporal choices.

No Evidence for Influence of Demand Characteristics.
To ascertain that EFT constitutes a viable option for affecting intertemporal choices, it is critical to understand whether the observed effects may simply reflect demand characteristics (Orne, 1962). Indeed, it has been shown that, under certain circumstances, participants can deduce the experimenter's intention from the instructions (Rung & Madden, 2018; Stein et al., 2018) - though accounting for such insight did not alter the results (Stein et al., 2018).

Notably, a recent study compared EFT with a control condition that was carefully constructed to induce a greater demand effect (Rung & Madden, 2019). Yet, the EFT condition did show a stronger beneficial impact on intertemporal choices. EFT has also been shown to alter choices in ecologically more valid settings that are presumably less susceptible to demand effects. For example, such studies assessed, as dependent measures, calories intake in a food court (O'Neill et al., 2016) and real-life weight loss (Sze et al., 2015). These considerations are consistent with our analysis: The studies that were *least likely* prone to demand effects showed in fact the greatest effect.

Caveats.
EFT may not always be adaptive. Though exaggerated discounting is clearly maladaptive (Amlung et al., 2019), there are several reasons why it is not always beneficial to nudge decisions via EFT. First, given the inherent uncertainty of the future, choices focusing on the immediate present are not always myopic (Hayden, 2016). It is, for example, sometimes better to choose a smaller but certain reward over a larger one that may not actually materialize (Bulley et al., 2016).

Second, our analysis indicates that the impact of EFT is also based on an "affective forecast" of the delayed choice option (Benoit et al., 2019; Boyer, 2008). However, these forecasts are often not very accurate. For example, there is a tendency to overestimate the intensity and duration of positive reward-related emotions (Bulley et al., 2017; Wilson & Gilbert, 2005), suggesting that we may, at times, overvalue the future.

Third, the laboratory tasks indicate that EFT provides motivational incentives for choices that can immediately be acted upon (i.e., by foregoing an immediate but smaller reward). However, long lasting behavior change may additionally require the identification of possible obstacles (Oettingen & Gollwitzer, 2018; Oettingen & Reininger, 2016). This research indeed indicates that merely fantasizing about a positive future event can make it less likely that people then work towards achieving that goal.

Finally, not all people may equally profit from EFT. Although our data show that individuals with greater choice impulsivity benefit from this mechanism, this may not be the case for people with emotional disorders such as anxiety or depression. They often find it more difficult to conjure positive imaginings of the future (Gamble et al., 2019), suggesting that they would also struggle in creating a positive mental image of the delayed reward (but see also Ji et al., 2017; Renner et al., 2017).

Publication Bias. A major strength of this meta-analysis is its analytical approach. In our three-level model, we comprehensively included all effect sizes, thus avoiding selection bias (Cheung, 2019), and accounted for dependency with robust variance estimation (Hedges et al., 2010). In addition, we made efforts to account for publication bias. We carefully searched for unpublished experiments, resulting in 16% of effect sizes stemming from such sources. We further employed various methods to adjust for remaining publication bias, given that there are issues with virtually any one of them (Carter et al., 2019; Renkewitz & Keiner, 2019). Though they were somewhat inconclusive with regards to the question whether the study pool is still affected by bias, all adjustments revealed a significant and stable effect size estimate. In the absence of any preregistered, large-scale replication studies, we thus suggest that this meta-analysis currently provides the best available evidence for the impact of EFT on intertemporal choices.

Conclusions. The current meta-analysis provides a comprehensive and quantitative synthesis of EFT and its influence on farsighted decisions. Our moderator analyses highlight critical candidate core components of EFT and suggest avenues for future research aimed at delineating their interactions and unique contributions. By showing that EFT is particularly effective in nudging individuals characterized by choice impulsivity, we corroborate that it may be a helpful mechanism for optimizing future-oriented decisions. We further envision that EFT may have a similar impact in other domains that inherently entail a tradeoff with the future, including societal challenges such as promoting more sustainable behavior (Bø & Wolff, 2020; Lee et al., 2020; Williams & Benoit, 2021).

Context. Following the seminal contribution by Peters & Büchel, 2010 and our subsequent publication (Benoit et al., 2011), we were excited to witness the accumulation of research into the impact of EFT on intertemporal choice. A particular important development was the generalization of this work from the delay discounting of monetary rewards to the examination of health-relevant choices (Daniel et al., 2013). Given the progress of this field and its translational potential (e.g., Stein et al., 2016), we found it prudent to take stock of the extant literature. In particular, we saw the opportunity to exploit the heterogeneity of the experimental protocols across studies to identify critical core components of EFT. We thus hope that the current meta-analysis does not only corroborate the impact of EFT on intertemporal choice, but also informs future research into the mechanisms by which EFT promotes more farsighted decisions.

ACKNOWLEDGEMENTS

All authors were funded by a Max Planck Research Group awarded to RGB. We thank Vera Seyffert, Janine Held and Anna Stanzel for assistance during the literature search and variable coding, Martina Dietrich for help with formatting the manuscript, Angharad Williams for comments on a previous draft, and Wolfgang Viechtbauer for advice on data analysis.

Author note: All data and analysis scripts are publicly available at the *Open Science Framework*: https://osf.io/9rejf/?view_only=fee62b3cadd4405b8aafdac75015f60b
Sarah A. Rösch https://orcid.org/0000-0001-5439-8147
Davide F. Stramaccia https://orcid.org/0000-0002-9744-0445
Roland G. Benoit https://orcid.org/0000-0002-7127-4899
The authors have no known conflict of interest to disclose.

References

MAd: Meta-analysis with mean differences.

Addis, D. R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology*.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. & Csaki, B. F., editors, *Second international symposium on information theory*, pages 267–281. Academiai Kiado, Budapest. Publisher: Akadémiai Kiadó, Budapest.

Amasino, D. R., Sullivan, N. J., Kranton, R. E., & Huettel, S. A. (2019). Amount and time exert independent influences on intertemporal choice. *Nature Human Behaviour*, 3(4):383–392.

Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., Naish, K. R., Reed, D. D., & McCabe, R. E. (2019). Delay discounting as a transdiagnostic process in psychiatric disorders: A meta-analysis. *JAMA Psychiatry*, 76(11):1176.

Amlung, M., Petker, T., Jackson, J., Balodis, I., & MacKillop, J. (2016). Steep discounting of delayed monetary and food rewards in obesity: a meta-analysis. *Psychological Medicine*, 46(11):2423–2434.

Amlung, M., Vedelago, L., Acker, J., Balodis, I., & MacKillop, J. (2017). Steep delay discounting and addictive behavior: a meta-analysis of continuous associations: Delay discounting and addiction. *Addiction*, 112(1):51–62.

Appelhans, B. M., Woolf, K., Pagoto, S. L., Schneider, K. L., Whited, M. C., & Liebman, R. (2011). Inhibiting food reward: Delay discounting, food reward sensitivity, and palatable food intake in overweight and obese women. *Obesity*, 19(11):2175–2182.

Assink, M. & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12(3):154–174.

Atance, C. M. & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences*, 5(12):533–539.

Banes, K. E. (2016). *The effect of episodic future thought on delay discounting, outcome expectancies, and alcohol use among risky college drinkers* [Doctoral dissertation, Virginia Polytechnic Institute and State University]. ProQuest Dissertations and Theses Global.

Baumeister, R. F. & Masicampo, E. J. (2010). Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal–culture interface. *Psychological Review*, 117(3):945–971.

Baumgartner, H., Pieters, R., & Bagozzi, R. P. (2008). Future-oriented emotions: conceptualization and behavioral effects. *European Journal of Social Psychology*, 38(4):685–696.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41(2):257–278.

Benoit, R. G., Berkers, R. M. W. J., & Paulus, P. C. (2018). An adaptive function of mental time travel: Motivating farsighted decisions. *The Behavioral and brain sciences*, 41:e3. Place: England.

Benoit, R. G., Gilbert, S. J., & Burgess, P. W. (2011). A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(18):6771–6779.

Benoit, R. G., Paulus, P. C., & Schacter, D. L. (2019). Forming attitudes via neural activity supporting affective episodic simulations. *Nature Communications*, 10(1):2215.

Berns, G. S., Laibson, D., & Loewenstein, G. (2007). Intertemporal choice – toward an integrative framework. *Trends in Cognitive Sciences*, 11(11):482–488.

Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: delay discounting in current, never, and ex-smokers. *Psychopharmacology*, 146(4):447–454.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. John Wiley & Sons, West Sussex, UK.

Boyer, P. (2008). Evolutionary economics of mental time travel? *Trends in Cognitive Sciences*, 12(6):219–224.

Bromberg, U., Wiehler, A., & Peters, J. (2015). Episodic future thinking is related to impulsive decision making in healthy adolescents. *Child Development*, 86(5):1458–1468.

Bulley, A., Henry, J., & Suddendorf, T. (2016). Prospection and the present moment: The role of episodic foresight in intertemporal choices between immediate and delayed rewards. *Review of General Psychology*, 20(1):29–47.

Bulley, A., Henry, J. D., & Suddendorf, T. (2017). Thinking about threats: Memory and prospection in human threat management. *Consciousness and Cognition*, 49:53–69.

Bulley, A., Miloyan, B., Pepper, G. V., Gullo, M. J., Henry, J. D., & Suddendorf, T. (2019). Cuing both positive and negative episodic foresight reduces delay discounting but does not affect risk-taking. *Quarterly Journal of Experimental Psychology*, 72(8):1998–2017.

Bulley, A. & Schacter, D. L. (2020). Deliberating trade-offs with the future. *Nature Human Behaviour*, 4(3):238–247.

Bø, S. & Wolff, K. (2020). I can see clearly now: Episodic future thinking and imaginability in perceptions of climate-related risk events. *Frontiers in Psychology*, 11:218.

Calluso, C., Tosoni, A., Cannito, L., & Committeri, G. (2019). Concreteness and emotional valence of episodic future thinking (EFT) independently affect the dynamics of intertemporal decisions. *PLOS ONE*, 14(5):e0217224. Publisher: Public Library of Science.

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: a comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, 2(2):115–144.

Cheng, Y.-Y., Shein, P. P., & Chiou, W.-B. (2012). Escaping the impulse to immediate gratification: The prospect concept promotes a future-oriented mindset, prompting an inclination towards delayed gratification: Prospect and delayed gratification. *British Journal of Psychology*, 103(1):129–141.

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19(2):211–229.

Cheung, M. W.-L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, 29(4):387–396.

Chiou, W.-B. & W.-H. Wu (2017). Episodic future thinking involving the nonsmoking self can induce lower discounting and cigarette consumption. *Journal of Studies on Alcohol and Drugs*, 78(1):106–112.

Ciaramelli, E., Sellitto, M., Tosarelli, G., & di Pellegrino, G. (2019). Imagining events alternative to the present can attenuate delay discounting. *Frontiers in Behavioral Neuroscience*, 13:269.

Cooper, N. (2013). *The neural representation of value and individual differences in human intertemporal choice* (UMI no. 3609138) [Doctoral dissertation, University of Pennsylvania]. ProQuest Dissertations and Theses Global.

Dai, X. & Fishbach, A. (2013). When waiting to choose increases patience. *Organizational Behavior and Human Decision Processes*, 121(2):256–266.

Daniel, T. O., Sawyer, A., Dong, Y., Bickel, W. K., & Epstein, L. H. (2016). Remembering versus imagining: when does episodic retrospection and episodic prospection aid decision making. *Journal of Applied Research in Memory and Cognition*, 5(3):352–358.

Daniel, T. O., Stanton, C. M., & Epstein, L. H. (2013). The future is now: Comparing the effect of episodic future thinking on impulsivity in lean and obese individuals. *Psychological Science*, 24(11):2339–2342.

D'Argembeau, A. & van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness and Cognition*, 13(4):844–858.

Dassen, F. C., Jansen, A., Nederkoorn, C., & Houben, K. (2016). Focus on the future: Episodic future thinking reduces discount rate and snacking. *Appetite*, 96:327–332.

Daugherty, J. R. & Brase, G. L. (2010). Taking time to be healthy: Predicting health behaviors with delay discounting and time perspective. *Personality and Individual Differences*, 48(2):202–207. Place: Netherlands Publisher: Elsevier Science.

De Luca, F., Benuzzi, F., Bertossi, E., Braghittoni, D., di Pellegrino, G., & Ciaramelli, E. (2018). Episodic future thinking and future-based decision-making in a case of retrograde amnesia. *Neuropsychologia*, 110:92–103.

de Wit, H. (2002). Acute administration of d-amphetamine decreases impulsivity in healthy volunteers. *Neuropsychopharmacology*, 27(5):813–825.

Dreves, P. A. & Blackhart, G. C. (2019). Thinking into the future: how a future time perspective improves self-control. *Personality and Individual Differences*, 149:141–151.

Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2013). How decisions emerge: Action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General*, 142(1):93–100.

Duval, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2):455–463.

D'Argembeau, A. & Van der Linden, M. (2012). Predicting the phenomenology of episodic future thoughts. *Consciousness and Cognition*, 21(3):1198–1206.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109):629–634.

Epstein, L. H., Richards, J. B., Saad, F. G., Paluch, R. A., Roemmich, J. N., & Lerman, C. (2003). Comparison between two measures of delay discounting in smokers. *Experimental and Clinical Psychopharmacology*, 11(2):131–138. Place: US Publisher: American Psychological Association.

Ernst, A., Philippe, F., & d'Argembeau, A. (2018). Wanting or having to: The role of goal self-concordance in episodic future thinking. *Consciousness and Cognition*, 66(6):26–39.

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2019). Detecting selection bias in meta-analyses with multiple outcomes: A simulation study. *The Journal of Experimental Education*, pages 1–20. Publisher: Routledge.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.

Gamble, B., Tippett, L. J., Moreau, D., & Addis, D. R. (2019). The futures we want: How goal-directed imagination relates to mental health. preprint, PsyArXiv.

Gilbert, D. T. & Wilson, T. D. (2007). Prospection: Experiencing the future. *Science*, 317(5843):1351–1354.

Green, L. & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic reward. *Psychological Bulletin*, 130(5):769–792.

Green, L., Myerson, J., & Macaux, E. W. (2005). Temporal discounting when the choice is between two delayed reward. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5):1121–1133.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, 83(2):314–320.

Gregory, W. L., Cialdini, R. B., & Carpenter, K. M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so? *Journal of Personality and Social Psychology*, 43(1):89–99. Place: US Publisher: American Psychological Association.

Hall, P. A., Fong, G. T., & Sansone, G. (2015). Time perspective as a predictor of healthy behaviors and disease-mediating states. *Time perspective theory; Review, research and application: Essays in honor of Philip G. Zimbardo.*, pages 339–352. Place: Cham, Switzerland Publisher: Springer International Publishing.

Hamilton, K. R., Mitchell, M. R., Wing, V. C., Balodis, I. M., Bickel, W. K., Fillmore, M., Lane, S. D., Lejuez, C. W., Littlefield, A. K., Luijten, M., Mathias, C. W., Mitchell, S. H., Napier, T. C., Reynolds, B., Schütz, C. G., Setlow, B., Sher, K. J., Swann, A. C., Tedford, S. E., White, M. J., Winstanley, C. A., Yi, R., Potenza, M. N., & Moeller, F. G. (2015). Choice impulsivity: Definitions, measurement issues, and clinical implications. *Personality Disorders: Theory, Research, and Treatment*, 6(2):182–198.

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). *dmetar: Companion R package for the guide 'Doing meta-analysis in R* (0.0.9000) [Computer software]. https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_R/dmetar.html.

Hayden, B. Y. (2016). Time discounting and time preference in animals: A critical review. *Psychonomic Bulletin & Review*, 23(1):39–53.

Hedges, L. V. & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9(4):426–445.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1):39–65.

Hendrickson, K. L. & Rasmussen, E. B. (2013). Effects of mindful eating training on delay and probability discounting for food and money in obese and healthy-weight individuals. *Behaviour Research and Therapy*, 51(7):399–409.

Higgins, J. D. & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011].* The Cochrane Collaboration.

Hollis-Hansen, K., Seidman, J., O'Donnell, S., & Epstein, L. H. (2019). Episodic future thinking and grocery shopping online. *Appetite*, 133:1–9.

Holmes, E. A. & Mathews, A. (2010). Mental imagery in emotion and emotional disorders. *Clinical Psychology Review*, 30(3):349–362.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Multilevel analysis: Techniques and applications, 2nd ed. Routledge/Taylor & Francis Group, New York, NY, US, 2nd ed. edition.

Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis: correcting error and bias in research findings*. Sage Publications, Newbury Park.

J. Q. Wu, Szpunar, K. K., Godovich, S. A., Schacter, D. L., & Hofmann, S. G. (2015). Episodic future thinking in generalized anxiety disorder. *Journal of Anxiety Disorders*, 36:1–8.

Ji, J. L., Holmes, E. A., & Blackwell, S. E. (2017). Seeing light at the end of the tunnel: Positive prospective mental imagery and optimism in depression. *Psychiatry Research*, 247:155–162.

Joireman, J., Balliet, D., Sprott, D., Spangenberg, E., & Schultz, J. (2008). Consideration of future consequences, ego-depletion, and self-control: Support for distinguishing between CFC-Immediate and CFC-Future sub-scales. *Personality and Individual Differences*, 45(1):15–21.

Knapp, G. & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17):2693–2710.

Kollins, S. H. (2003). Delay discounting is associated with substance use in college students. *Addictive Behaviors*, 28(6):1167–1173.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis: Three-level meta-analysis. *Research Synthesis Methods*, 2(1):61–76.

Kromrey, J. D. & Rendina-Gobioff, G. (2006). On knowing what we do not know: An empirical comparison of methods to detect publication bias in meta-analysis. *Educational and Psychological Measurement*, 66(3):357–373. Publisher: SAGE Publications Inc.

Kurth-Nelson, Z., Bickel, W., & Redish, A. D. (2012). A theoretical account of cognitive effects in delay discounting. *European Journal of Neuroscience*, 35(7):1052–1064.

Kwan, D., Craver, C. F., Green, L., Myerson, J., Boyer, P., & Rosenbaum, R. S. (2012). Future decision-making without episodic mental time travel. *Hippocampus*, 22(6):1215–1219.

Kwan, D., Craver, C. F., Green, L., Myerson, J., Gao, F., Black, S. E., & Rosenbaum, R. S. (2015). Cueing the personal future to reduce discounting in intertemporal choice: Is episodic prospection necessary? *Hippocampus*, 25(4):432–443.

Kwan, D., Craver, C. F., Green, L., Myerson, J., & Rosenbaum, R. S. (2013). Dissociations in future thinking following hippocampal damage: evidence from discounting and time perspective in episodic amnesia. *Journal of experimental psychology. General*, 142(4):1355–1369. Place: United States.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4.

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *BMJ*, 333(7568):597–600.

Lebreton, M., Bertoux, M., Boutet, C., Lehericy, S., Dubois, B., Fossati, P., & Pessiglione, M. (2013). A Critical Role for the Hippocampus in the Valuation of Imagined Outcomes. *PLoS Biology*, 11(10):e1001684.

Lee, P.-S., Sung, Y.-H., Wu, C.-C., Ho, L.-C., & Chiou, W.-B. (2020). Using episodic future thinking to pre-experience climate change increases pro-environmental behavior. *Environment and Behavior*, 52(1):60–81.

Lempert, K. M., Johnson, E., & Phelps, E. A. (2016). Emotional arousal predicts intertemporal choice. *Emotion*, 16(5):647–656.

Lempert, K. M. & Phelps, E. A. (2016). The malleability of intertemporal choice. *Trends in Cognitive Sciences*, 20(1):64–74.

Levens, S. M., Sagui-Henson, S. J., Padro, M., Martin, L. E., Trucco, E. M., Cooperman, N. A., Baldwin, A. S., Kassianos, A. P., & Mdege, N. D. (2019). The effects of positive affect and episodic future thinking on temporal discounting and healthy food demand and choice among overweight and obese individuals: Protocol for a pilot 2×2 factorial randomized controlled study. *JMIR research protocols*, 8(3):e12265–e12265. Publisher: JMIR Publications.

Lin, H. & Epstein, L. H. (2014). Living in the moment: Effects of time perspective and emotional valence of episodic thinking on delay discounting. *Behavioral Neuroscience*, 128(1):12–19.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*, volume Thousand Oaks. Sage Publications, Calif.

Liu, L., Feng, T., Chen, J., & Li, H. (2013). The value of emotion: how does episodic prospection modulate delay discounting? *PloS one*, 8(11):e81717.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, 97(387):666.

Loewenstein, G. & Lerner, J. S. (2003). The role of affect in decision making. In Davidson, R., Goldsmith, H., & Scherer, K., editors, *Handbook of affective sciences*, pages 619–642. Oxford University Press.

López-López, J. A., Van den Noortgate, W., Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2017). Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: A monte carlo simulation. *Research Synthesis Methods*, 8(4):435–450.

Lüdecke, D. (2019). *Esc: effect size computation for meta analysis* (0.5.1.) [Computer software]. https://doi.org/10.5281/zenodo.1249218.

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, 20(4):641–654.

MacKillop, J., Amlung, M. T., Few, L. R., Ray, L. A., Sweet, L. H., & Munafò, M. R. (2011). Delayed reward discounting and addictive behavior: a meta-analysis. *Psychopharmacology*, 216(3):305–321.

Martínez-Loredo, V., Fernández-Hermida, J. R., Carballo, J. L., & Fernández-Artamendi, S. (2017). Long-term reliability and stability of behavioral measures among adolescents: The delay discounting and stroop tasks. *Journal of Adolescence*, 58:33 – 39.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In Commons, M. L., Mazur, J. E., Nevin, J. A., & Rachlin, H., editors, *Quantitative analyses of behavior: The effect of delay and of intervening events on reinforcement value.*, volume 5, pages 55–73. Lawrence Erlbaum Associates, Inc. Place: Hillsdale, NJ, US Publisher: Lawrence Erlbaum Associates, Inc.

McCormick, C., Ciaramelli, E., De Luca, F., & Maguire, E. A. (2018). Comparing and Contrasting the Cognitive Effects of Hippocampal and Ventromedial Prefrontal Cortex Damage: A Review of Human Lesion Studies. *Neuroscience*, 374:295–318.

Mellis, A. M., Snider, S. E., Deshpande, H. U., LaConte, S. M., & Bickel, W. K. (2019). Practicing prospection promotes patience: Repeated episodic future thinking cumulatively reduces delay discounting. *Drug and Alcohol Dependence*, 204:107507.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, 6(7):e1000097. Publisher: Public Library of Science.

Morris, S. B. & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, 7(1):105–125.

Morrison, K. L., Madden, G. J., Odum, A. L., Friedel, J. E., & Twohig, M. P. (2014). Altering impulsive decision making with an acceptance-based procedure. *Behavior Therapy*, 45(5):630–639.

O'Donnell, S., Oluyomi Daniel, T., & Epstein, L. H. (2017). Does goal relevant episodic future thinking amplify the effect on delay discounting? *Consciousness and Cognition*, 51:10–16.

Oettingen, G. & Gollwitzer, P. M. (2018). Health behavior change by self-regulation of goal pursuit. In de Ridder, D., Adriaanse, M., & Fujita, K., editors, *The Routledge international handbook of self-control in health and well-being*, pages 418–430. Routledge, New York.

Oettingen, G. & Reininger, K. M. (2016). The power of prospection: mental contrasting and behavior change: Mental contrasting and behavior change. *Social and Personality Psychology Compass*, 10(11):591–604.

O'Neill, J., Daniel, T. O., & Epstein, L. H. (2016). Episodic future thinking reduces eating in a food court. *Eating Behaviors*, 20:9–13.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11):776–783.

Palombo, D., Keane, M., & Verfaellie, M. (2015). The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking. *Hippocampus*, 25(3):345–353.

Palombo, D., Keane, M., & Verfaellie, M. (2016). Using future thinking to reduce temporal discounting: Under what circumstances are the medial temporal lobes critical? *Neuropsychologia*, 89:437–444.

Paolacci, G. & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3):184–188.

Paret, C., Jennen-Steinmetz, C., & Schmahl, C. (2017). Disadvantageous decision-making in borderline personality disorder: Partial support from a meta-analytic review. *Neuroscience & Biobehavioral Reviews*, 72:301–309.

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences*, 19(10):590–602.

Peters, J. & Büchel, C. (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, 66(1):138–148.

Peters, J. & Büchel, C. (2011). The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends in Cognitive Sciences*, 15(5):227–239.

R Core Team (2019). *R: a language and environment for statistical computing* (3.6.1) [Computer software]. R Foundation for Statistical Computing.

Race, E., Keane, M. M., & Verfaellie, M. (2011). Medial temporal lobe damage causes deficits in episodic memory and episodic future thinking not attributable to deficits in narrative construction. *Journal of Neuroscience*, 31(28):10262–10269.

Read, D., Frederick, S., Orsel, B., & Rahman, J. (2005). Four score and seven years from now: The date/delay effect in temporal discounting. *Management Science*, 51(9):1326–1335.

Renkewitz, F. & Keiner, M. (2019). How to detect publication bias in psychological research: A comparative evaluation of six statistical methods. *Zeitschrift für Psychologie*, 227(4):261–279.

Renner, F., Ji, J. L., Pictet, A., Holmes, E. A., & Blackwell, S. E. (2017). Effects of engaging in repeated mental imagery of future positive events on behavioural activation in individuals with major depressive disorder. *Cognitive Therapy and Research*, 41(3):369–380.

Rick, S. & Loewenstein, G. (2008). The role of emotion in economic behavior. In Lewis, M., Haviland-Jones, J. M., & Barrett, L. F., editors, *Handbook of emotions*, pages 138–156. The Guilford Press, New York, NY, US, 3 edition.

Rohatgi, A. (2019). *WebPlotDigitizer* (4.2.) [Computer software]. https://automeris.io/WebPlotDigitizer.

Rollins, B. Y., Dearing, K. K., & Epstein, L. H. (2010). Delay discounting moderates the effect of food reinforcement on energy intake among non-obese women. *Appetite*, 55(3):420–425.

Rothstein, H., Sutton, A. J., & Borenstein, M., editors (2005). *Publication bias in meta-analysis: prevention, assessment and adjustments*. Wiley, Chichester, England ; Hoboken, NJ. OCLC: ocm61211258.

Rung, J. M. & Madden, G. J. (2018). Experimental reductions of delay discounting and impulsive choice: A systematic review and meta-analysis. *Journal of Experimental Psychology: General*, 147(9):1349–1381.

Rung, J. M. & Madden, G. J. (2019). Demand characteristics in episodic future thinking II: The role of cues and cue content in changing delay discounting. *Experimental and Clinical Psychopharmacology*, 27(5):482–495.

Schacter, D. L., Benoit, R. G., De Brigard, F., & Szpunar, K. K. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of Learning and Memory*, 117:14–21.

Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic future thinking: mechanisms and functions. *Current Opinion in Behavioral Sciences*, 17:41–50.

Schubert, T., Eloo, R., Scharfen, J., & Morina, N. (2020). How imagining personal future scenarios influences affect: Systematic review and meta-analysis. *Clinical Psychology Review*, 75:101811.

Sheffer, C. E., Mackillop, J., Fernandez, A., Christensen, D., Bickel, W. K., Johnson, M. W., Panissidi, L., Pittman, J., Franck, C. T., Williams, J., & Mathew, M. (2016). Initial examination of priming tasks to decrease delay discounting. *Behavioural Processes*, 128:144–152.

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality and Social Psychology Bulletin*, 11(1):118–127. Publisher: SAGE Publications Inc.

Shevorykin, A., Pittman, J. C., Bickel, W. K., O'connor, R. J., Malhotra, R., Prashad, N., & Sheffer, C. E. (2019). Primed for health: Future thinking priming decreases delay discounting. *Health Behavior and Policy Review*, 6(4):363–377.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2):534–547.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General*, 144(6):1146–1152.

Stein, J. S., Tegge, A. N., Turner, J. K., & Bickel, W. K. (2018). Episodic future thinking reduces delay discounting and cigarette demand: an investigation of the good-subject effect. *Journal of behavioral medicine*, 41(2):269–276. Place: United States.

Stein, J. S., Wilson, A. G., Koffarnus, M. N., Daniel, T. O., Epstein, L. H., & Bickel, W. K. (2016). Unstuck in time: episodic future thinking reduces delay discounting and cigarette smoking. *Psychopharmacology*, 233(21-22):3771–3778.

Steinberg, L., Graham, S., O'Brien, L., Woolard, J., Cauffman, E., & Banich, M. (2009). Age differences in future orientation and delay discounting. *Child development*, 80(1):28–44. Place: United States.

Story, G. W., Vlaev, I., Seymour, B., Darzi, A., & Dolan, R. J. (2014). Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Frontiers in Behavioral Neuroscience*, 8.

Strickland, J. C., Lile, J. A., & Stoops, W. W. (2019). Evaluating non-medical prescription opioid demand using commodity purchase tasks: test-retest reliability and incremental validity. *Psychopharmacology*, 236(9):2641–2652.

Suddendorf, T. & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, 30(3):299–313.

Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *BMJ*, 320(7249):1574–1577.

Sze, Y. Y., Daniel, T. O., Kilanowski, C. K., Collins, R. L., & Epstein, L. H. (2015). Web-based and mobile delivery of an episodic future thinking intervention for overweight and obese families: A feasibility study. *JMIR mHealth and uHealth*, 3(4):e97.

Szpunar, K. K. & Schacter, D. L. (2013). Get real: effects of repeated simulation and emotion on the perceived plausibility of future experiences. *Journal of experimental psychology. General*, 142(2):323–327.

Szpunar, K. K., Spreng, R. N., & Schacter, D. L. (2014). A taxonomy of prospection: Introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences*, 111(52):18414–18421.

Tanner-Smith, E. E. & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and SPSS: Robust variance estimation. *Research Synthesis Methods*, 5(1):13–30.

Thorstad, R. & Wolff, P. (2018). A big data analysis of the relationship between future thinking and decision-making. *Proceedings of the National Academy of Sciences*, 115(8):E1740–E1748.

Trope, Y. & Liberman, N. (2010). "Construal-level theory of psychological distance": Correction to Trope and Liberman (2010). *Psychological Review*, 117(3):1024–1024.

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses based on p-values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science*, 11(5):713–729.

van Aert, R. M. C. (2019). *puniform: Meta-analysis methods correcting for publication bias* (0.2.1) [Computer software]. https://CRAN.R-project.org/package=puniform.

van Aert, R. M. C. & van Assen, M. A. L. M. v. A. (2020). Correcting for publication bias in a meta-analysis with the p-uniform* method. Technical report, OSF.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3).

Wagenmakers, E.-J. & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1):192–196.

Weafer, J., Baggott, M. J., & de Wit, H. (2013). Test–retest reliability of behavioral measures of impulsive choice, impulsive action, and inattention. *Experimental and Clinical Psychopharmacology*, 21(6):475–481.

Williams, A. N. & Benoit, R. G. (2021). Imagining our better selves. *Nat Sustain*.

Wilson, T. D. & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, 14(3):131–134.

Wu, W.-H., Cheng, W., & Chiou, W.-B. (2017). Episodic future thinking about the ideal self induces lower discounting, leading to a decreased tendency toward cheating. *Frontiers in Psychology*, 8.

Yi, R. (2020). *CAPT*. Department of Psychology, University of Kansas.

Zhang, S., Peng, J., Qin, L., Suo, T., & Feng, T. (2018). Prospective emotion enables episodic prospection to shift time preference. *British Journal of Psychology*, 109(3):487–499. Publisher: John Wiley & Sons, Ltd.

Table 1. Moderator Coding Criteria

Moderator and levels	Criteria
Choice domain	
Monetary	Any measure of monetary delay discounting (e.g., AUC, *k*-values)
Health-relevant	Either food (e.g., energy intake), smoking (e.g., demand intensity), or alcohol-related measures (e.g., demand intensity)
Valence	
Positive Positive-to-neutral Neutral Negative	Assignment was based on ratings if available. Otherwise, assignment was based on descriptions. Only conditions that were explicitly deemed positive were assigned to this level. When it was ambiguous whether all episodes were necessarily positive (i.e., some may have been neutral), they were assigned to the positive-to-neutral level.
Vividness	Either vividness or imagery score or any composite score of these scores. The standardized mean difference or standardized mean change between the EFT and control group was derived as a continuous moderator.
Episodicity	
Control: others	All other control conditions.
Control: Semantic future	The control task also required participants to consider the future but in a non-episodic fashion.
Future-orientedness	
Control: others	All other control conditions.
Control: non future, episodic	The control task also entailed simulations of episode that, however, did not take place in the possible future.

Table 1 continued

Moderator and levels	Criteria
Content specificity	
Content-specific	Any episode that involved spending the reward (i.e., the money) Any episode that included behaviors instrumental to achieve a health-related goal or that entailed the consequences of its achievement
Non content-specific	Any general type of EFT unrelated to the reward
Choice impulsivity	
Individuals with choice impulsivity	Individuals that were overweight or obese and those with an addiction to nicotine or other substances
Neurotypical individuals	All other samples
Design	
Between-subjects	Effect size estimates from independent groups
Within-subjects	Effect size estimates from repeated measures
Proneness to Demand effects [a]	
Most likely	- Within-subjects designs: participants were informed about the concrete study purpose; between-subject designs: participants in the EFT but not in the control groups were informed that the study was about farsighted decisions (no equal bias for both groups). - No “active” control task (e.g. baseline condition only) - Article reported high expectancy rates for EFT group.
Intermittent	- Participants were not informed about the concrete study purpose, but performed both a EFT and control condition.
Least likely	- Study involved a control condition that made it difficult to discern the study purpose (e.g. episodic recent thinking; semantic future thinking). - Participants had to self-generate episodic cues in both the EFT and control condition. - Participants did not know about the existence of a control condition in between-subjects design studies.

Moderator and levels	Criteria
Publication status	
Unpublished	Dissertations, unpublished data, in-prep manuscripts
Published	Peer-reviewed articles
Study site	
On-site	Any study conducted in person (mostly in-lab)
Online	Any study not conducted in person, e.g. via crowdsourcing platforms such as MTurk

[a] Cohen's κ was run to determine if there was agreement between the two raters' judgements on the probability that the study purpose was correctly discerned. There was high agreement between the two raters, $\kappa = .92$, $z = 15.5$, $p < .001$.

Table 2. Sample and Moderator Information for Studies Included in the Meta-Analysis

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1001	m	EFT familiar vs. Control	pos	n/a	0	0	0	n	w	0	1	1	22	-0.06	0.04
1001	m	EFT unfamiliar vs. Control	pos	n/a	0	0	0	n	w	0	1	1	22	0.00	0.04
1002	m	EFT vs. SFT	pos	n/a	1	0	0	n	b	0	3	1	60	0.66	0.07
1002	m	EFT vs. Control	pos	n/a	0	0	0	n	b	0	3	1	60	0.83	0.07
1002	m	EFT vs. Present self	pos	n/a	0	1	0	n	b	0	3	1	60	0.71	0.07
1002	m	EFT vs. Control	pos	n/a	0	0	0	n	b	0	3	1	60	0.64	0.07
1003	m	Goal: EFT vs. ERT	pos	3.94	0	1	1	n	b	0	3	1	52	10.94	0.09
1003	m	General: EFT vs. ERT	pos	4.92	0	1	0	n	b	0	3	1	52	0.67	0.08
1004	m	EFT vs. SFT	pos	n/a	1	0	0	y	b	0	3	1	60	0.57	0.07
1004	m	EFT vs. Control	pos	n/a	0	0	0	y	b	0	3	1	60	0.63	0.07
1004	h	EFT vs. SFT	pos	n/a	1	0	1	y	b	0	3	1	60	0.74	0.07
1004	h	EFT vs. Control	pos	n/a	0	0	1	y	b	0	3	1	60	0.82	0.07
1005	m	EFT vs. Control	pos	0.82	0	0	0	n	w	0	2	1	37	0.64	0.03
1005	h	Intensity: EFT vs. Control	pos	0.82	0	0	0	n	w	0	2	1	37	0.30	0.02
1005	h	Breakpoint: EFT vs. Control	pos	0.82	0	0	0	n	w	0	2	1	37	-0.06	0.02

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1005	h	Omax: EFT vs. Control	pos	0.82	0	0	0	n	w	0	2	1	37	0.39	0.02
1005	h	Pmax: EFT vs. Control	pos	0.82	0	0	0	n	w	0	2	1	37	0.65	0.03
1005	h	Elasticity: EFT vs. Control	pos	0.82	0	0	0	n	w	0	2	1	37	0.00	0.02
1007	m	EFT vs. ERT	pos	n/a	0	1	0	y	b	0	3	1	42	0.67	0.10
1007	h	EFT vs. ERT	pos	n/a	0	1	0	y	b	0	3	1	41	0.71	0.10
1009	m	EFT vs. ERT	pos	n/a	0	1	0	y	b	0	3	1	50	0.94	0.09
1009	h	EFT vs. ERT	pos	n/a	0	1	0	y	b	0	3	1	37	2.19	0.17
1009	h	EFT vs. ERT	pos	n/a	0	1	0	y	b	0	3	1	37	-0.28	0.11
1012	m	EFT general vs. EPT general	pos	-0.29	0	1	0	n	b	0	3	1	47	0.50	0.09
1012	m	EFT food vs. EPT food	pos	-0.05	0	1	0	n	b	0	3	1	47	0.47	0.09
1012	h	EFT general vs. EPT general	pos	-0.29	0	1	0	n	b	0	3	1	47	-0.13	0.09
1012	h	EFT food vs. EPT food	pos	-0.05	0	1	1	n	b	0	3	1	47	0.73	0.09
1016	m	EFT familiar vs. Control	pos	n/a	0	0	0	n	w	0	1	1	23	0.26	0.04
1016	m	EFT unfamiliar vs. Control	pos	n/a	0	0	0	n	w	0	1	1	23	0.27	0.04
1019	m	EFT vs. ERT	pos	n/a	0	1	0	n	b	0	3	1	42	1.06	0.11
1019	h	EFT vs. ERT	pos	n/a	0	1	0	n	b	0	3	1	42	0.27	0.10
1024	m	EFT vs. Baseline	pos	n/a	1	0	1	n	w	0	1	1	13	1.26	0.12

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1028	m	EFT vs. EPT	p/n	n/a	0	1	0	n	b	0	3	1	87	0.45	0.05
1029	m	EFT vs. Control	pos	n/a	0	0	0	n	w	0	1	1	32	0.51	0.03
1029	m	EFT vs. Control	neg	n/a	0	0	0	n	w	0	1	1	31	-0.46	0.03
1029	m	EFT vs. Control	neut	n/a	0	0	0	n	w	0	1	1	30	0.01	0.03
1031	m	AUC $10: EFT vs. CET	pos	0.66	0	1	0	n	w	0	2	1	48	0.23	0.02
1031	m	AUC $100: EFT vs. CET	pos	0.66	0	1	0	n	w	0	2	1	48	0.30	0.02
1034	m	EFT vs. EPT	neut	n/a	0	1	0	n	b	0	3	1	64	0.63	0.07
1037	m	EFT vs. Estimate	p/n	n/a	1	0	1	n	w	0	2	1	12	0.30	0.07
1037	m	EFT vs. Estimate	p/n	n/a	1	0	1	n	w	0	2	1	12	0.27	0.07
1041	m	EFT vs. Control	p/n	n/a	0	0	0	n	w	0	1	1	46	0.51	0.02
1048	m	EFT vs. Control	p/n	n/a	0	0	0	n	w	0	1	1	32	0.28	0.03
1048	m	EFT vs. Control	p/n	n/a	0	0	0	y	w	0	1	1	30	0.54	0.03
1050	m	1 Cue adjusting amount: EFT 1 vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	64	0.26	0.06
1050	m	3 Cue adjusting amount: EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	67	0.59	0.06
1050	m	1 Cue delay: EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	64	0.15	0.06

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1050	m	3 Cue delay: EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	67	0.82	0.06
1051	m	EFT vs. Control	pos	n/a	0	0	0	n	w	0	1	1	34	0.91	0.04
1051	m	EFT vs. Control	pos	n/a	0	0	0	n	w	0	1	1	34	0.58	0.03
1051	m	EFT vs. Baseline	pos	n/a	0	0	0	n	w	0	1	1	34	0.87	0.03
1051	m	EFT vs. Control	neut	n/a	0	0	0	n	w	0	1	1	34	0.09	0.02
1051	m	EFT vs. Control	neut	n/a	0	0	0	n	w	0	1	1	34	0.05	0.02
1051	m	EFT vs. Baseline	neut	n/a	0	0	0	n	w	0	1	1	34	-0.18	0.02
1051	m	EFT vs. Control	neg	n/a	0	0	0	n	w	0	1	1	32	-1.00	0.04
1051	m	EFT vs. Control	neg	n/a	0	0	0	n	w	0	1	1	32	-0.76	0.03
1051	m	EFT vs. Baseline	neg	n/a	0	0	0	n	w	0	1	1	32	-0.68	0.03
1052	m	EFT vs. Control	p/n	n/a	0	0	0	n	w	0	1	1	44	0.78	0.03
1054	m	EFT vs. ERT	pos	0.05	0	1	0	y	b	1	3	1	66	0.92	0.07
1054	m	EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	136	0.67	0.03
1054	m	EFT vs. No ET	pos	n/a	0	0	0	y	b	1	3	1	136	0.61	0.03
1054	h	Omax: EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	136	0.53	0.03
1054	h	Elasticity: EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	136	0.53	0.03

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1054	h	Breakpoint: EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	136	0.34	0.03
1054	h	Intensity: EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	136	0.31	0.03
1054	h	Pmax: EFT vs. ERT	pos	n/a	0	1	0	y	b	1	3	1	136	-0.16	0.03
1054	h	Omax: EFT vs. No ET	pos	n/a	0	0	0	y	b	1	3	1	136	0.97	0.03
1054	h	Elasticity: EFT vs. No ET	pos	n/a	0	0	0	y	b	1	3	1	136	0.90	0.03
1054	h	Breakpoint: EFT vs. No ET	pos	n/a	0	0	0	y	b	1	3	1	136	0.31	0.03
1054	h	Intensity: EFT vs. No ET	pos	n/a	0	0	0	y	b	1	3	1	136	0.80	0.03
1054	h	Pmax: EFT vs. No ET	pos	n/a	0	0	0	y	b	1	3	1	136	-0.16	0.03
1056	m	EFT vs. ERT	pos	0.25	0	1	0	y	b	1	1	1	117	0.37	0.03
1056	h	Demand intensity: EFT vs. ERT	pos	0.25	0	1	0	y	b	1	1	1	113	0.40	0.04

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1056	h	Elasticity: EFT vs. ERT	pos	0.25	0	1	0	y	b	1	1	1	117	0.15	0.03
1056	h	Craving: EFT vs. ERT	pos	0.25	0	1	0	y	b	1	1	1	117	0.16	0.03
1072	m	EFT vs. Present	neut	n/a	0	0	0	n	w	0	2	1	21	0.47	0.04
1072	m	EFT vs. Present	neut	n/a	0	0	0	n	w	0	2	1	21	0.10	0.04
1072	m	EFT vs. Present	neut	n/a	0	0	0	n	w	0	2	1	22	-0.35	0.04
1072	m	EFT vs. Present	neut	n/a	0	0	0	n	w	0	2	1	25	0.09	0.03
1077	m	EFT vs. CRT	pos	0.64	0	1	0	n	b	0	3	1	54	0.44	0.08
1077	m	EFT vs. EPT	pos	0.1	0	1	0	n	b	0	3	1	54	0.22	0.07
1102	m	EFT change vs. AAT or Control change	pos	n/a	0	0	0	y	b	0	3	1	60	0.39	0.08
1102	h	EFT: pre- vs. post-training	pos	n/a	0	0	1	y	w	0	3	1	20	-0.16	0.04
1103	m	AUC $10: EFT vs. Control	pos	1.18	0	1	0	y	b	0	3	1	26	1.42	0.19
1103	m	AUC $100: EFT vs. Control	pos	1.18	0	1	0	y	b	0	3	1	26	1.45	0.20
1103	h	EFT vs. Control energy intake	pos	1.18	0	1	0	y	b	0	3	1	26	1.05	0.18

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1104	h	Energy: EFT child vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	0.68	0.21
1104	h	Red Foods: EFT child vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	0.30	0.20
1104	h	Green Foods: EFT child vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	0.00	0.20
1104	h	BMI reduction: EFT child vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	0.35	0.21
1104	h	% overweight: EFT child vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	0.32	0.20
1104	m	AUC: EFT child vs. Control	pos	n/a	0	0	0	y	b	0	3	0	20	-0.38	0.21
1104	h	Ad libitum energy intake: EFT child vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	-0.25	0.20
1104	h	Energy: EFT parent vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	0.97	0.23
1104	h	Red Foods: EFT parent vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	0.22	0.20
1104	h	Green Foods: EFT parent vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	0.61	0.21

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1104	h	BMI reduction: EFT parent vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	1.53	0.26
1104	h	% overweight: EFT parent vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	1.60	0.27
1104	m	AUC: EFT parent vs. Control	pos	n/a	0	0	0	y	b	0	3	0	20	0.75	0.22
1104	h	Ad libitum energy intake: EFT parent vs. Control	pos	n/a	0	0	1	y	b	0	3	1	20	1.32	0.25
1105	h	Calories: EFT vs. ERT	pos	0.34	0	1	1	y	b	0	3	1	29	1.17	0.16
1105	h	% calories from fat: EFT vs. ERT	pos	0.34	0	1	1	y	b	0	3	1	29	1.25	0.17
1105	h	% calories from protein: EFT vs. ERT	pos	0.34	0	1	1	y	b	0	3	1	29	1.11	0.16
1105	h	% calories from carbs: EFT vs. ERT	pos	0.34	0	1	1	y	b	0	3	1	29	-0.38	0.14
1110	m	EFT vs. Control	pos	0	0	1	0	n	b	0	3	1	200	0.63	0.02
1110	m	EFT vs. Control	pos	0	0	1	0	n	b	0	3	1	200	0.62	0.02
1110	m	EFT vs. Control	pos	0	0	1	0	n	b	0	3	1	200	0.39	0.02
1110	m	EFT vs. Control	neg	-0.24	0	1	0	n	b	0	3	1	198	0.52	0.02

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1110	m	EFT vs. Control	neg	-0.24	0	1	0	n	b	0	3	1	198	0.46	0.02
1110	m	EFT vs. Control	neg	-0.24	0	1	0	n	b	0	3	1	198	0.39	0.02
1111	m	EFT vs. ERT	pos	0.11	0	1	0	n	b	0	3	1	36	0.58	0.12
1111	m	EFT vs. SET	pos	0.19	0	1	0	n	b	0	3	1	35	0.94	0.13
1113	h	EFT health vs. Savings Control	pos	2.90	0	1	1	y	b	0	3	1	24	1.98	0.25
1113	h	EFT health vs. Savings Control	pos	2.90	0	1	1	y	b	0	3	1	24	1.49	0.21
1113	h	EFT health vs. Savings Control	pos	2.90	0	1	1	y	b	0	3	1	24	0.59	0.18
1113	h	EFT health vs. Savings Control	pos	2.90	0	1	1	y	b	0	3	1	24	1.39	0.21
1113	h	EFT general vs. ERT	pos	0.96	0	1	0	y	b	0	3	1	23	0.89	0.19
1113	h	EFT health vs. ERT	pos	2.16	0	1	1	y	b	0	3	1	21	0.99	0.21
1113	h	EFT general vs. ERT	pos	0.96	0	1	0	y	b	0	3	1	23	0.27	0.18
1113	h	EFT health vs. ERT	pos	2.16	0	1	1	y	b	0	3	1	21	0.65	0.20
1113	h	EFT general vs. ERT	pos	0.96	0	1	0	y	b	0	3	1	23	0.02	0.17
1113	h	EFT health vs. ERT	pos	2.16	0	1	1	y	b	0	3	1	21	-0.25	0.19
1113	h	EFT general vs. ERT	pos	0.96	0	1	0	y	b	0	3	1	23	1.24	0.21
1113	h	EFT health vs. ERT	pos	2.16	0	1	1	y	b	0	3	1	21	1.18	0.22
1114	m	EFT vs. ERT	pos	n/a	0	1	0	n	b	1	3	1	160	0.46	0.03
1116	m	EFT vs. Baseline	pos	n/a	0	0	0	n	w	0	1	1	55	0.54	0.02
1116	m	EFT vs. Baseline	neut	n/a	0	0	0	n	w	0	1	1	55	0.52	0.02
1116	m	EFT vs. Baseline	neg	n/a	0	0	0	n	w	0	1	1	55	0.25	0.02

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1118	m	Episodic vs. Control	pos	n/a	0	0	0	n	b	0	2	0	60	1.67	0.09
1145	m	k Visit 1: EFT vs. ERT	pos	0.27	0	1	0	n	b	0	3	0	33	-0.50	0.13
1145	m	AUC Visit 1: EFT vs. ERT	pos	0.27	0	1	0	n	b	0	3	0	33	-0.38	0.12
1145	h	Calories Visit 1: EFT vs. ERT	pos	0.27	0	1	0	n	b	0	3	0	33	-0.07	0.12
1150	m	Patient prospective: After vs. Before	neut	n/a	0	0	1	n	w	0	1	0	8	1.67	0.28
1169	m	Cued 6 months: EFT vs. ERT	pos	n/a	0	1	0	n	b	1	3	1	176	0.53	0.02
1169	m	Uncued 6 months: EFT vs. ERT	pos	n/a	0	1	0	n	b	1	3	1	174	0.00	0.02
1169	m	Cued 1 year: EFT vs. ERT	pos	n/a	0	1	0	n	b	1	3	1	176	0.90	0.03
1169	m	Uncued 1 year: EFT vs. ERT	pos	n/a	0	1	0	n	b	1	3	1	174	0.19	0.02
1169	m	EFT-typical vs. ERT-time	pos	n/a	0	1	0	n	b	1	3	1	145	0.51	0.03
1169	m	EFT-event vs. ERT-time	pos	n/a	0	1	0	n	b	1	3	1	137	0.63	0.03
1169	m	EFT-uncued vs. ERT-time	pos	n/a	0	1	0	n	b	1	3	1	136	-0.13	0.03

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1170	m	EFT vs. Control	neut	n/a	0	0	0	n	w	0	3	1	24	0.10	0.03
1170	m	EFT vs. Control	neut	n/a	0	0	0	n	w	0	3	1	20	0.11	0.04
1170	m	EFT vs. Control	neut	n/a	0	0	0	n	w	0	3	1	24	0.00	0.03
1170	m	EFT vs. Control	neut	n/a	0	0	0	n	w	0	3	1	20	0.02	0.04
1171	m	EFT vs. Baseline	pos	n/a	0	0	0	n	w	0	3	1	22	0.16	0.04
1171	m	EFT vs. Baseline	pos	n/a	0	0	0	n	w	0	3	1	22	0.18	0.04
1171	m	EFT vs. Baseline	pos	n/a	0	0	0	n	w	0	3	1	22	0.29	0.04
1193	m	Process: EFT vs. ERT	pos	n/a	0	1	0	n	b	0	3	1	42	1.25	0.11
1193	m	No Process: EFT vs. ERT	pos	n/a	0	1	0	n	b	0	3	1	36	1.03	0.13
1195	m	EFT vs. Baseline	pos	n/a	0	0	0	n	w	0	1	1	50	0.34	0.02
1207	m	EFT vs. EPT	pos	-0.43	0	1	0	n	b	0	1	0	85	0.08	0.05
1207	m	EFT vs. CET	pos	0.72	0	0	0	n	b	0	1	0	86	0.25	0.05
1207	h	Alcohol use intention: EFT vs. EPT	pos	-0.43	0	1	0	n	b	0	1	0	85	0.09	0.05
1207	h	Demand intensity: EFT vs. EPT	pos	0.72	0	1	0	n	b	0	1	0	85	0.36	0.05
1207	h	Alcohol use intention: EFT vs. CET	pos	-0.43	0	0	0	n	b	0	1	0	86	0.19	0.05

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1207	h	Demand intensity: EFT vs. CET	pos	0.72	0	0	0	n	b	0	1	0	86	0.38	0.05
1221	m	EFT money vs. Past money	p/n	-0.2	0	1	0	n	b	0	2	1	115	-0.06	0.03
1221	m	EFT money vs. Present imagine	p/n	-0.33	0	0	0	n	b	0	2	1	131	0.06	0.03
1221	m	EFT money vs. Present attend	p/n	0	0	0	0	n	b	0	2	1	122	0.28	0.03
1235	m	EFT pre vs. EFT post	pos	n/a	0	0	0	n	w	0	1	0	32	1.46	0.06
1240	h	Intensity: EFT vs. Present	neut	n/a	0	1	0	n	b	0	2	0	80	-0.36	0.05
1240	h	Breakpoint: EFT vs. Present	neut	n/a	0	1	0	n	b	0	2	0	80	-0.09	0.05
1240	h	Omax: EFT vs. Present	neut	n/a	0	1	0	n	b	0	2	0	80	-0.19	0.05
1240	h	Pmax: EFT vs. Present	neut	n/a	0	1	0	n	b	0	2	0	80	-0.05	0.05
1240	h	Elasticity: EFT vs. Present	neut	n/a	0	1	0	n	b	0	2	0	80	-0.31	0.05
1240	h	Intensity: EFT vs. Control	neut	n/a	0	0	0	n	w	0	2	0	40	0.03	0.02
1240	h	Breakpoint: EFT vs. Control	neut	n/a	0	0	0	n	w	0	2	0	40	0.04	0.02

Table 2 continued

StudyID	Choice Domain	Condition	Theoretically-Motivated Moderators						Procedural Moderators				n	Effect Size	Sampling Variance
			Valence	Vividness	Episodicity[a]	Future-orientedness[b]	Content Specificity	Choice Impulsivity	Design	Online?	Proneness to Demand Effects[c]	Published?			
1240	h	Omax: EFT vs. Control	neut	n/a	0	0	0	n	w	0	2	0	40	-0.03	0.02
1240	h	Pmax: EFT vs. Control	neut	n/a	0	0	0	n	w	0	2	0	40	0.02	0.02
1240	h	Elasticity: EFT vs. Control	neut	n/a	0	0	0	n	w	0	2	0	40	0.00	0.02
1241	m	positive EFT vs. neutral	pos	n/a	0	0	0	n	w	0	2	0	35	0.17	0.02
1241	m	Certain future vs. Control	neut	n/a	0	0	0	n	w	0	2	0	32	-0.07	0.03
1241	m	Positive future pre vs. post	pos	n/a	0	0	0	n	w	1	2	0	64	-0.04	0.01
1330	m	EFT vs. Baseline	p/n	n/a	0	0	0	n	w	0	1	1	20	0.84	0.06

Note. y = yes (in the case of choice impulsivity. e.g., smokers), applicable; n = no, non-applicable; AAT = Approach Avoidance Task; b = between-subjects; BMI = Body Mass Index; CET = Control Episodic Thinking; AUC = area-under-the-curve; CRT = Control Recent Thinking; EFT = Episodic Future Thinking; EPT = Episodic Past Thinking; ERT = Episodic Recent Thinking; h = health-relevant; m = monetary; neg = negative; neut = neutral; n/a = not available; No ET = No Episodic Thinking; Omax = maximum expenditure; Pmax = the price point corresponding to Omax; p/n = positive-to-neutral; pos = positive; SET = Standardized Episodic Thinking; SFT = Semantic Future Thinking; w = within-subjects. Effect size is reported as Hedges' *g*. Associations between StudyID and full references are provided in Supplemental Table S1.

[a] 0 =Control: Others, 1 = Control: Semantic. [b] 0 =Control: Others, 1 = Control: Non. [c] 1 =most likely, 2 = intermittent, 3 = least likely.

Table 3. Moderator Analyses

Moderator and level	ES	s	*g*	95% CI	*SE*	*F*-value[a]	*p*-value[a]	*t*-value[b]	*p*-value[b]	Level 2 variance	Level 3 variance	Q_E(df)	*p*-value	I^2
Choice domain						$F(1, 46) = 0.98$	.328			.081	.082	681.75 (172)	< .001	79.49 [72.73, 85.49]
Monetary	103	54	0.462	[0.343, 0.58]	0.06			7.83	< .001					
Health-relevant	71	18	0.385	[0.231, 0.538]	0.08			5.05	< .001					
Valence [c]						$F(3, 44) = 3.02$	.04			.061	.064	555.73 (170)	< .001	74.64 [65.88 82.49]
Negative	8	4	-0.178	[-0.877, 0.518]	0.33			-0.518	.607					
Neutral	26	10	0.184	[-0.018, 0.385]	0.1			1.83	.074					
Positive-to-neutral	11	7	0.438	[0.236, 0.639]	0.1			4.37	< 0.001					
Positive	129	42	0.52	[0.4, 0.639]	0.06			8.79	< .001					
Vividness	60	15				$F(1, 13) = 8.43$	.012	2.91[d]	.012[d]	.013	.069	126.51 (58)	< .001	62.99 [34.55, 83.55]
Episodicity						$F(1, 46) = 0.774$	.384			.084	.076	678.38 (172)	< .001	79.23 [72.47, 85.27]

Table 3 continued

Moderator and level	ES	s	g	95% CI	SE	F-value[a]	p-value[a]	t-value[b]	p-value[b]	Level 2 variance	Level 3 variance	Q_E(df)	p-value	I^2
Control: others	168	56	0.433	[0.324, 0.542]	0.054			8.0	< .001					
Control: Semantic future	6	4	0.553	[0.285, 0.821]	0.133			4.15	< .001					
Future-orientedness						$F(1, 46) = 0.07$	.789			.084	.076	663.1 (172)	< .001	79.22 [72.34; 85.4]
Control: others	90	35	0.43	[0.307, 0.553]	0.061			6.51	< .001					
Control: non future, episodic	84	28	0.45	[0.312, 0.587]	0.068			6.57	< .001					
Content specificity						$F(1, 46) = 9.21$	.004			.084	.064	657.19 (172)	< .001	77.9 [70.89, 84.23]
Non content-specific	141	53	0.391	[0.285, 0.821]	0.13			7.46	< .001					
Content-specific	33	12	0.707	[0.502, 0.912]	0.11			6.95	< .001					
Choice impulsivity						$F(1, 46) = 6.23$	.016			.086	.059	639.37 (172)	< .001	77.51 [70.36, 84.03]
Neurotypical	108	44	0.367	[0.246, 0.487]	0.060			6.12	< .001					

Table 3 continued

Moderator and level	ES	s	*g*	95% CI	*SE*	*F*-value[a]	*p*-value[a]	*t*-value[b]	*p*-value[b]	Level 2 variance	Level 3 variance	Q_E(df)	*p*-value	I^2
Choice impulsivity	66	15	0.604	[0.452, 0.755]	0.075			8.02	< .001					
Design						*F*(1, 46) = 3.8	.057			.088	.055	621.71 (172)	< .001	77.19 [69.97, 83.75]
Between-subject	116	33	0.509	[0.368, 0.651]	0.07			7.26	< .001					
Within-subjects	58	27	0.319	[0.188, 0.45]	0.065			4.89	< .001					
Study site						*F*(1, 46) = 1.2	.278			.082	.081	678.74 (172)	< .001	79.56 [77.84, 85.54]
On-site	144	50	0.451	[0.33, 0.572]	0.06			7.50	<.001					
Online	30	8	0.377	[0.285, 0.469]	0.046			8.27	< .001					
Proneness to Demand effects[d]						*F*(2, 45) = 3.97	.026			.087	.051	597.01 (171)	< .001	76.62 [69.06, 83.53]
Most likely	38	17	0.381	[0.187, 0.574]	0.096			3.96	< .001					

Table 3 continued

Moderator and level	ES	s	*g*	95% CI	*SE*	*F*-value[a]	*p*-value[a]	*t*-value[b]	*p*-value[b]	Level 2 variance	Level 3 variance	Q_E(df)	*p*-value	I^2
Intermittent	31	11	0.201	[-0.005, 0.407]	0.102			1.97	.055					
Least likely	105	30	0.532	[0.406, 0.658]	.063			8.5	< .001					
Publication status						*F*(1, 46) = 0.55	.463			.085	.071	639.16 (172)	< .001	78.72 [71.71, 85.02]
Unpublished	27	11	0.307	[-0.096, 0.71]	0.2			1.53	0.132					
Published	147	49	0.457	[0.359, 0.556]	0.049			9.37	< .001					

Note. ES = number of effect size estimates; *s* = number of studies; *F*-value = test of moderator significance; *g* = Hedges' standardized difference on each level of the moderator; 95% CI correspond to the CIs of Hedges' *g* on each level, *p* corresponds to the *p*-value for the omnibus test for the *F*-value or to the significance on the level of the moderator, respectively. The number of effect size estimates and studies often do not add up as expected because some studies provided multiple effect size estimates and/or did not provide data for the level of a moderator.

[a] Omnibus test. [b] Test for the coefficient. [c] Follow-up-analyses showed that positive episodes have a significantly stronger beneficial effect as compared to neutral episodes, β = 0.336, $t(44) = 2.96$, $p = .005$, and a marginally more beneficial effect relative to negative episodes, β = 0.699, $t(44) = 1.94$, $p = .059$. Positive-to-neutral episodes were only associated with a marginally greater effect than neutral, β = 0.254, $t(44) = 1.8$, $p = .08$, and negative episodes, β = 0.617, $t(44) = 1.71$, $p = .094$. None of the other direct comparisons was significant or showed a trend, all $t(44) < 1.29$, all $p > .2$. [d] Test for coefficient of an increase in 1 SD in vividness. [e] Follow-up-analyses showed that studies where participants were least likely to discern the study purpose showed a greater effect size than the ones with an intermittent likelihood, β = 0.331, $t(45) = 2.76$, $p = .008$. Other differences between the moderator levels were not significant.

Table 4. Results for the Multiple Moderator Model

Moderator variables	β (SE)	95% CI	*p*-value	*t*-value
Intercept	0.201 (0.13)	[-0.058, 0.478]	.121	1.58
Negative episode (vs. neutral)	-0.404 (0.28)	[-0.97, 0.163]	.158	-1.44
Positive-to-neutral episode (vs. neutral)	0.250 (0.13)	[-0.017, 517]	.055	1.89
Positive episode (vs. neutral)	0.265 (0.13)	[-0.001, 0.53]	.051	2.02
Content-specific episodes (vs. non content-specific)	0.245 (0.11)	[0.017, 0.473]	.036	2.17
Neurotypical sample (vs. sample characterized by choice impulsivity)	0.015 (0.11)	[-0.203, 0.233]	.89	0.14
Discerning study purpose intermittent (vs. least likely)	-0.192 (0.14)	[-0.472, 0.088]	.173	-1.39
Discerning study purpose most likely (vs. least likely)	-0.042 (0.13)	[-0.214, 0.298]	.74	0.33
Omnibus test		*F*(7, 40) = 3.68, *p* = .004		
Variance within studies		0.062		
Variance between studies		0.053		
Number of ESs		174		

Note. CI = Confidence Interval, ES = Effect Size.