Supporting Information for
Exposing Omitted Moderators: Explaining Differences in Treatment Effects in the Social Sciences.

Antonia Krefeld-Schwalb [*a], Eli Rosen Sugerman [b], and Eric Johnson [b]
[a] Rotterdam School of Management, Erasmus University, [b] Columbia Business School, Columbia University

[*]Antonia Krefeld-Schwalb
Email: krefeldSchwalb@rsm.nl

1 Additional Material for Study 1 and 2

5.5 **Experimental Design.** To study heterogeneity across the panels, we directly replicated five behavioral effects. In selecting behavioral effects for inclusion, we sought effects demonstrated to be robust, amenable to online survey, and relevant to behavioral marketing. Additionally, we sought a wide range of effect sizes, reasoning that smaller effects may be more sensitive to panel differences.

To ensure robustness, we selected only effects that were either successfully replicated in one of the Many Labs replication projects or qualified by successful meta-analysis. The sunk cost effect, replicated by Many Labs (Klein et al. 2014), describes individuals' increased likelihood to follow through with a venture when they have invested more resources into it. The framing effect, replicated by Many Labs (Klein et al. 2014), describes the influence of gain versus loss framing on individuals' willingness to take risks (we used the unusual disease task). The less-is-better effect, replicated by Many Labs (Klein et al. 2018), describes individuals' propensity to value a less expensive gift as more generous than a more expensive gift. The default effect, supported by meta-analysis (Jachimowicz et al. 2019), describes the increased likelihood for individuals to choose the pre-selected default option (we use Johnson and Goldstein's (2013) organ donation paradigm). We also included the local warming effect, supported by meta-analysis (Sugerman et al. 2021) describes the influence of recent, local temperature on perceptions of climate change (we use the two-question paradigm from Li et al. 2011). For local warming, we asked for respondents' zip code in order to gather the weather and historic climate data through the API of the National Center for Environmental Information. Since this is not a randomized experiment, we did not include it in the following analyses. All effects can be tested with one or two questions and don't require long readings or otherwise extensive effort from the respondents, making them particularly well-suited to be tested in one, all-encompassing online survey. We followed the procedures documented in the replication studies. In addition to typical demographic variables such as gender, age, income, political orientation, and education; we also asked how many hours respondents spend on the respective survey platforms (except in the student sample), we collected cognitive measures of *Fluid Intelligence* (cognitive reflection task; Berlin numeracy task), and we included two attention checks questions. Acknowledging the distinct categories of attention check type questions noted by Kung et al. (2018), we use both one instructed-response item (wherein the question tells subjects how to respond) and one instructional manipulation check (wherein the correct answer is typically implied and obfuscated within the text of the question). In recognition of Hauser & Schwarz' (2016) caution that MTurk is a population that *learns*, and Thomas & Clifford's (2017) contention that MTurk subjects are often aware of formulaic attention checks, we rely on novel attention check items. Finally, we follow the guidance of Oppenheimer et al. (2009) to align the attention check questions to the rest of the questionnaire, using one shorter and one longer, vignette-like question. Our first check states "I typically work twenty-eight hours in a day." On the 7-point Likert scale of agreement, attentive respondents note "Disagreement" or "Strong Disagreement." Our second check, drafted for this

project, reads: “Many states collect a tax on the sale of various goods and services, called sales tax. Imagine that your friend sends you to the grocery store with $1. He asks you to purchase 9 apples, 6 pears, 3 oranges, and a pineapple. The pineapple costs $2.00, the apples are on sale for $.25/each, and the pears and oranges cost $.75/each. As you wait in line to pay for your produce, another customer hands you a coupon for 9 free apples. Please select “Sales tax affects local economies” from the options below.” Attentive respondents select the answer that reads “Sales tax affects local economies” from the six multiple choice options.

1.2 Study 1a. In Study 1a (preregistered at doi.org/1.17605/OSF.IO/4DYP2) we distributed one questionnaire across 8 unique panels. Three of these panels were selected for their relevance to academic research: Amazon’s Mechanical Turk (MTurk), Cloud Research (the MTurk Toolkit product), and Prolific. For the remaining panels, we relied on Lucid Marketplace, which is a large panel aggregator. We include Lucid’s default product, which is a “blended” sample drawing indiscriminately from any of their suppliers, as well as four specific panel suppliers available on the Lucid Marketplace. In order to increase the likelihood of panel variation, we used two panels from the Lucid Marketplace that Lucid’s own quality scoring indicates were low-quality (inBrain.ai and Tap Research) and two indicated as high-quality (Branded Research and Prodege). We used two of Qualtrics’ native algorithms to detect likely bots, which were excluded from further analysis. The sample sizes before and after filtering for bots are reported in Table S1.

Upon the initial soft-launch of the survey and through subsequent conversations with Lucid Marketplace representatives, we learned that some panel suppliers avoid sending panelists to surveys when they are the first or only supplier on the project. For this reason, we reselected suppliers, following the same empirical methodology as described above, but restricting our set to those suppliers which Lucid representatives indicated are okay to allow themselves to be the first supplier on a given project. We also added a panel to our analysis, which is Lucid Marketplace’s typical panel offering (an indiscriminate blend from all suppliers). One implication of this adjustment is that respondents were recruited in consecutive batches. It is possible that the latter batches may differ significantly from the earlier ones with respect to moderator variables. Thus, we analyzed the data for differences in these respective variables both across batches and using a survey date measure (the timestamped date of survey completion) and found no difference.

1.3 Study 1b. Study 1b was a preregistered extension of Study 1a (doi.org/1.17605/OSF.IO/HGDQK) using Prolific’s “representative” United States and United Kingdom samples, as well as a (Dutch) undergraduate student sample. In order to adapt the survey to the UK and the Netherlands, respectively, the questions about political orientation were dropped and dollar amounts were converted to local currencies. Students were tested on computer in individual cubicles in the campus behavioral lab, which eliminated the need for bot detection in that sample.

Table S1

Sample sizes per panel with and without bots.

Study 1A			
Panel	N (Including Bots)	**N (Final)**	Compensation (per Subject)
MTurk	569	**453**	$1.00 USD
Cloud Research (MTurk Toolkit)	771	**740**	$1.00 USD
Prolific	509	**494**	$2.00 USD
Lucid Marketplace (Blend)	845	**837**	$1.50 USD
Branded Research	812	**802**	$1.50 USD

Prodege	502	**496**	$1.50 USD
inBrain.ai	538	**524**	$1.50 USD
TapResearch	502	**494**	$1.50 USD
Study 1B			
Students	605	**605**	Course Credit
Prolific UK representative	500	**498**	$2.00 USD
Prolific US representative	499	**496**	$2.00 USD
Study 2			
MTurk	485	**459**	$1.00 USD
Prolific	603	**598**	$2.00 USD
Study 3			
MTurk	600	**598**	$2.00 USD
Prolific	600	**598**	$3.00 USD
Study 4			
MTurk	480	**438**	$ 2.00 USD
Prolific	505	**498**	$ 3.00 USD
Lucid Marketplace (Blend)	500	**496**	$ 2.50 USD
Study 5			
MTurk	510	**468**	$ 2.00 USD
Prolific	504	**494**	$ 3.00 USD
Lucid Marketplace (Blend)	496	**494**	$ 2.50 USD

Table S2 – Descriptive for Demographics (After Bot Exclusion)

Study 1A

Panel	**MTurk**		**Prolific**		**Lucid (All) ****	
	Mean	SD	Mean	SD	Mean	SD
Age	39	10	37	13	48	17
Years of Education	16	1	15	2	14	2
Income*	$50,000 to $59,999		$50,000 to $59,999		$50,000 to $59,999	
Gender	62% Female; 34% Male; 4% Minority		67% Female; 30% Male; 3% Minority		64% Female; 35% Male; 1% Minority	

Study 1B

Panel	**Students**		**Prolific UK**		**Prolific US**	
Age	19	2	46	16	46	16
Years of Education	12	<1	15	2	15	2
Income	$70,000 to $79,999		£40,000 to £49,999		$50,000 to $59,999	
Gender	46% Female; 53% Male; 1% Minority		50% Female; 48% Male; 2% Minority		51% Female; 46% Male; 3% Minority	

Study 2

	MTurk		Prolific	
Age	38	10	38	13
EDU	16	2	15	2
Income	$50,000 to $59,999		$50,000 to $59,999	
Gender	41% Female; 59% Male;		62% Female; 34% Male; 4% Minority	

Study 3

Age	37	11	37	13
Years of Education	15	2	15	2
Income	$40,000 to $49,999		$40,000 to $49,999*	
Gender	36% Female; 63% Male; 1% Minority		50% Female; 45% Male; 5% Minority	

Study 4

	MTurk		Prolific		Lucid (Blended)	
Age	35	11	37	13	48	17
Years of Education	16	1	15	2	14	2
Income	$50,000 to $59,999*		$50,000 to $59,999		$30,000 to $39,999	
Gender	41% Female; 58% Male; 1% Minority		58% Female; 38% Male; 4% Minority		63% Female; 36% Male; 1% Minority	

Study 5

Age	37	10	41	14	50	18
Years of Education	16	2	15	2	14	2
Income	$50,000 to $59,999		$60,000 to $69,999		$40,000 to $49,999	
Gender	58% Female; 41% Male; 1% Minority		56% Female; 42% Male; 2% Minority		66% Female; 33% Male; 1% Minority	

Note: EDU is converted into the total number of years of education. * = Median 2020 Household Income Range. **For readability, the demographic information for Lucid (Blended) in Study 1A includes panelists from various sources recruited via Lucid such as Lucid Blended, Tap Research, and Prodege.

1.3 Study 2. Study 2 was a preregistered replication of Study 1 (doi.org/1.17605/OSF.IO/TKAQ2). We used the exact same survey as in Study 1, but the sole intention of the Study 1b was to replicate the reversal of effect sizes (i.e., the negative correlation) observed regarding the framing and default effects on MTurk and Prolific. We replicated the results of Study 1a. The framing effect had an effect size of d = 1.01 on Prolific, but .53 on MTurk; while the default effect showed the reverse pattern, with an effect of d = 1.03 on MTurk, but a .28 on Prolific.

Table S3

Overview of the paradigms tested in each of the studies.

		Study		
Paradigm	Description	1 and 2	3	4 and 5
Sunk Cost	The sunk cost effect describes individuals' increased likelihood to follow through with a venture when they have invested more resources into. Respondents are randomly assigned to either the condition with or without sunk costs. This paradigm has a two-condition, between-subjects experimental design. We follow the Oppenheimer et al. (2009) implementation of the paradigm. In both conditions, the dependent variable is a nine-point Likert scale ranging from "1 – Definitely stay at home" to "9 – Definitely go to the game" collected via multiple choice dropdown. Analysis focuses on mean differences.	X		X
Less-is-Better	The less-is-better effect describes individuals' propensity to value a less expensive gift as more generous than a more expensive. Respondents are randomly assigned to either the less or more expensive items conditions. This paradigm has a two-condition, between-subjects experimental design. We follow the Hsee (1998) implementation of the paradigm. In both conditions, the dependent variable is a seven-point Likert scale ranging from "0 – Not generous at all" to "6 – Extremely generous" collected via multiple choice dropdown. Analysis focuses on mean differences.	X		X
Framing	The framing effect describes the influence of gain versus loss framing on individuals' willingness to take risks. Respondents are randomly assigned to either the loss or gain frame condition. This paradigm has a two-condition, between-subjects experimental design. We follow the Tversky & Kahneman (1981) implementation of the paradigm. In both conditions, the dependent variable is respondents' choice between the two-items "Program A" and "Program B" collected via vertically aligned multiple choice. Analysis focuses on mean differences.	X		X
Default	The default effect describes the increased likelihood for individuals to choose the pre-selected default option. Respondents are randomly assigned to one of three conditions (opt-out, opt-in, or a neutral condition). This paradigm has a three-condition, between-subjects experimental design. We follow the Johnson & Goldstein (2003) implementation of the paradigm. In all conditions, the dependent variable is respondents' choice between two options, the default option or the changed status option. What varies across conditions is whether donor or non-donor status is defaulted. In the neutral condition, there is no default. Responses collected via horizontally aligned multiple choice. Analysis focuses on mean differences. In later studies, we omit the neutral condition.	X		X
Trolley Problem	The trolley problem regards ethical dilemmas surrounding the choice to sacrifice an individual to save a larger number of people. Respondents are randomly assigned to one of two		X	

	conditions. In each condition, respondents are presented with a two-option vertically aligned multiple choice question that asks for “Yes” or “No” determinations of moral permissibility. We follow the Hauser et al. (2007) implementation of the paradigm.		
Local Warming*	The local warming effect describes the influence of recent, local temperature on perceptions of climate change. This task is conducted in one condition only. It also utilizes actual temperature data as an instrumental variable. We follow the Li et al. (2010) implementation of the paradigm. The independent variable is a five-point Likert scale ranging from “Much warmer than usual” to “Much colder than usual” which is collected via a vertically aligned multiple choice question. The instrumental variable is objective temperature, which uses weather data associated with each respondent based on the day of survey completion and geographic location (zip code). The dependent variable is a four-point Likert scale ranging from “Not at all” to “A great deal” (regarding concern for global warming) and is collected via vertically aligned multiple choice question. Analysis uses either subjective temperature (the independent variable) or objective temperature (instrumental variable).	X	
False Consensus*	The false consensus effect describes the propensity for individuals to see their own choices as common and appropriate. There is only one condition. Three responses are collected. Numerical values are collected (via text entry questions) in response to two questions: 1) “What % of your peers do you estimate would sign the release?” and 2) “What % would refuse to sign it? (Total % should be 100%)”. The third question is a two-option, vertically aligned multiple choice between “Sign the release agreement” and “Refuse to sign the release agreement”. We follow the Ross et al. (1977) implementation of the paradigm.		X
Risk Preferences*	Risk preferences are measured in a titrator style, as adapted from Dohmen et al. (2011). In each item, respondents are asked to choose between “lottery” and “sure payout” in a vertically aligned multiple choice question. Analysis focuses on the correlation between choice based, quantitative methods and more qualitative self-report measures of economic preferences.		X
Time Preferences*	Time preferences are measured in a titrator style, as adapted from Dohmen et al. (2011). In each item, respondents are asked to choose between two payouts that vary in timeliness and dollar value, from a vertically aligned multiple choice question. Analysis focuses on the correlation between choice based, quantitative methods and more qualitative self-report measures of economic preferences.		X
Reciprocity*	Reciprocity is measured with five items, as adapted from Dohmen et al. (2011). Four of the items are 11-item Likert scales, asked in horizontally aligned multiple choice format. One item is a seven-item multiple choice question, aligned vertically. Analysis focuses on the correlation between choice based, quantitative methods and more qualitative self-report measures of economic preferences.		X

Note. For each paradigm tested in our studies, we replicated the original experiments as closely as possible given access to original study materials. For many of the paradigms, we accessed preregistered copies of the original survey questionnaires, allowing us to directly replicate most details. This includes the exact wording, the response style (e.g., multiple choice or text entry), and specific elements of the item design (e.g., multiple choice oriented vertically or horizontally; text entry force numeric response or no coded validation) of each of the original items. In Table S3 we describe key elements of each paradigm and indicate which studies included it. *The paradigms were included in the data collection but are not studied in this manuscript for the focus of the manuscript on text based interventions.

Table S4

Overview of respondents per condition in each study and panel

		Sunk Cost		Less is Better		Framing		Default		
	Condition	Base	Treat	Base	Treat	Base	Treat	Base	Neutral	Treat
Study	Panel									
1	Students	302	302	301	303	300	304	189	203	212
	Branded Research	393	409	392	410	400	402	269	268	265
	Prolific _UK_Rep	248	250	249	249	249	249	155	174	169
	Prolific	246	248	248	246	245	249	161	166	167
	Prolific _US_Rep	247	249	245	251	249	247	164	181	151
	Cloud Research	374	366	369	371	371	369	245	249	246
	Prodege	239	257	250	246	247	249	167	166	163
	Blended	433	404	412	425	410	427	259	273	305
	MTurk	228	225	226	227	227	226	150	151	152
	TapResearch	250	244	262	232	241	253	175	165	154
	inBrain.ai	261	263	252	272	270	254	175	176	173
2	Prolific	251	250	247	254	251	250	166	170	165
	MTurk	230	237	231	236	233	234	157	156	154
4	Prolific	250	248	248	250	246	252	165	167	166
	Lucid	249	247	250	246	256	240	152	172	172
	MTurk	219	219	223	215	218	220	145	151	142
5	Lucid	246	248	246	248	251	243	248	-	246
	Prolific	247	247	248	246	249	245	247	-	247
	MTurk	236	236	231	241	232	240	241	-	231
		Trolley								
		Base	Treat							
3	MTurk	299	299							
	Prolific	297	301							

2. Additional Results

2.1 Effect Sizes of the Behavioral Interventions

Table S5
Benchmark and average effect sizes of the behavioral interventions across all panels in Study 1 (a and b) and 3, 4, and 5.

	Study 1 (a and b)		
Effect	Benchmark	ES	CI
Sunk Cost (d)	.32a	.202	(.25,.15)
Less is Better (d)	.86b	.766	(.82,.72)
Framing Effect (h)	.58a	.584	(.54,.63)
Default Effect (out-in; h)	.68d	.478	(.42,.54)
	Study 2		
Sunk Cost (d)	.32a	.309	(.46,.16)
Less is Better (d)	.86b	.975	(1.13,.82)
Framing Effect (h)	.58a	.785	(.64,.93)
Default Effect (out-in; h)	.68d	.526	(.34,.71)
	Study 3		
Trolley Problem (h)	1.36b	1.014	(.88,1.15)
	Study 4		
Sunk Cost (d)	.32a	.242	(0.35,0.14)
Less is Better (d)	.86b	.740	(0.96,0.63)
Framing Effect (h)	.58a	.549	(0.44,0.65)
Default Effect (out-in; h)	.68d	.654	(0.53,0.78)
	Study 5		
Sunk Cost (d)	.32a	.179	(0.28,0.08)
Less is Better (d)	.86b	.636	(0.74,0.53)
Framing Effect (h)	.58a	.536	(0.43,0.64)
Default Effect (out-in; h)	.68d	.456	(0.35,0.56)

2.2 Testing Heterogeneity of Effect Sizes Across Panels. To test the heterogeneity of effects across studies, we implemented regression models predicting each paradigms' response distribution with dummy variables indicating the condition, the panel, and the interaction of condition and panel. We used logistic regression to model the binary responses in the framing and default paradigms. The interaction was used to test the heterogeneity of effect sizes across panels. As a measure of effect size, we next calculated the eta2 of the main effect of the conditions, panels, and interaction of the two factors. For significant interactions, we inferred that the effect sizes varied significantly between the panels. Table S5 reports the results of this analysis on the entire dataset of Study 1 (N = 6438) as well as the dataset filtered to include only those respondents who passed both attention check questions (N = 4037). All effect sizes, except for the local warming effect, varied significantly across panels. Local warming was also the only paradigm that we did not successfully replicate when aggregated across all panels.

Using seemingly unrelated regression (SUR) as an alternative approach to estimate the regression models resulted in the same inferences regarding the interaction of condition and panel.

Table S6
Effect sizes (η^2) for the effects of the manipulation (Cond), Panel and their interactions.

	Complete data (N = 6438)						Filtered data (N = 4037)		
	Without FACE			With FACE			Without FACE		
Effect	Cond	Panel	Panel* Cond	Cond	Panel	Panel* Cond	Cond	Panel	Panel* Cond
Sunk Cost	**.010*****	**.085*****	**.004****	**.010*****	**.045*****	**.003****	**.016*****	**.070*****	**.006****
Less is Better	**.128*****	**.054*****	**.007*****	**.127*****	**.019*****	**.006****	**.190*****	**.036*****	**.006*****
Framing Effect	**.061*****	**.013*****	**.008*****	**.062*****	**.003****	**.005****	**.089*****	**.006*****	**.004****
Default Effect	**.008*****	**.008*****	**.009*****	**.031*****	**.004*****	**.004****	**.009*****	**.009*****	**.006***

NOTE. The Table S reports the effects sizes in the complete as well as the filtered data (grey shaded cells) as well as in the absence (without FACE factors) and presence (with FACE factors) of factors and their interactions with the condition in the models. Significant effects are bold and the asterisk indicate the *p*-value. * $p < .05$, ** $p < .01$, *** $p < .05$

2.3 The FACE Factor Model. To organize the FACE factor variables, we ran an exploratory factor analysis with varimax rotation. The factor analysis revealed four factors that explained, in total, 47% of the variance in the FACE factor variables (see Table S3). The factors oblimin rotated, but the strongest correlation r = -.39. The demographic variables age, and income did not load highly (r > .3) onto any of the four factors, nor did they constitute their own factor (see Table S5).

The cognitive reflection task, numeracy, and attention checks loaded onto one factor that we termed *Fluid Intelligence*. Response time measures loaded onto one factor that we termed *Attentiveness*, representing the effort that respondents spent on the experimental tasks. Note that response times were logarithmized and scaled prior to entering the analysis.

Table S7

Eigenvalues and explained variance for factors of FACE factor variables

	Experience	Attentiveness	Crystallized Knowledge	Fluid Intelligence
Eigenvalues	2.985	2.804	1.246	1.14
Explained variance	0.187	0.175	0.078	0.071
Cumulative explained variance	0.187	0.362	0.44	0.511

Table S8
Cross-factor correlations

	Fluid Intelligence	Attentiveness	Crystallized Knowledge	Experience
Fluid Intelligence	1			
Attentiveness	.129	1		
Crystallized Knowledge	.078	-.115	1	
Experience	-.261	-.362	.061	1

Table S9
Factor loadings of FACE factor variables.

Study 1

	Factors			
	Fluid Intelligence	Attentiveness	Crystallized Knowledge	Experience
Cognitive Reflection Task (CRT) – Experience	**.769**	-.018	.036	-.008
Berlin Numeracy Task (BNT) – Experience	**.705**	.042	-.03	.009
Sunk Cost – RT	-.04	**.73**	.022	.036
Less is Better -RT	-.005	**.794**	-.011	-.01
Local Warming – Response time (RT)	-.048	**.614**	-.004	.165
Default – RT	.026	**.783**	-.022	-.089
Framing – RT	.064	**.743**	-.001	-.032
Age	-.243	.258	.201	-.132
Education	.003	-.006	**.999**	.007
Income	.121	.032	.284	.035
Attention Check	.247	.076	.142	-.276
Sunk Cost – Experience	-.023	-.007	.024	**.835**
Less is better – Experience	-.036	-.064	.028	**.8**
Local Warming – Experience	.035	.054	-.036	**.783**
Default – Experience	.072	-.039	.013	**.624**
Framing -Experience	-.01	.028	-.011	**.699**
	Study 4			
CRT	**0.667**	0.057	0.098	-0.057
BNT	**0.479**	0.116	0.049	0.081
Sunk Cost – RT	-0.017	**0.686**	0.014	0.084
Less is Better – RT	-0.020	**0.773**	-0.025	0.047

Default – RT	0.072	**0.754**	-0.050	-0.069
Framing – RT	0.015	**0.730**	0.076	-0.040
Synonym Task	0.000	-0.018	**0.856**	0.048
Antonym Task	0.059	0. 030	**0.784**	-0.086
Age	-0.378	0.183	**0.416**	-0.066
Education	0.128	-0.070	0.177	0.244
Income	0.250	-0.045	0.076	-0.019
Attention Check	0.384	0. 214	0.122	-0.238
Sunk Cost – Experience	-0.106	0.041	-0.076	**0.772**
Less is Better – Experience	-0.089	-0.037	-0.070	**0.714**
Default – Experience	-0.055	0.026	-0.015	**0.686**
Framing – Experience	0.044	0.039	0.035	**0.579**
	Study 5			
CRT	**0.544**	0.065	0.173	-0.057
Matrices	**0.403**	0.137	0.183	-0.076
3D-Rotation	**0.348**	-0.016	0.247	-0.048
Sunk Cost – RT	0.004	**0.743**	0.005	0.035
Less is Better – RT	-0.002	**0.84**	-0.05	0.037
Default – RT	-0.011	**0.793**	0.041	-0.015
Framing – RT	0.029	**0.765**	-0.002	-0.053
Synonym Task	0.014	0.013	**0.866**	0.039
Antonym Task	0.069	-0.003	**0.781**	-0.073
Age	-0.455	0.123	**0.428**	-0.09
Education	0.238	-0.114	0.174	0.242
Income	0.316	-0.05	0.143	0.032
Attention Check	0.354	0.168	0.08	-0.185
Sunk Cost – Experience	-0.017	0.01	-0.023	**0.834**
Less is Better – Experience	0.032	-0.021	0.086	**0.705**
Default – Experience	-0.008	-0.006	-0.009	**0.872**
Framing – Experience	-0.004	0.018	-0.013	**0.825**

NOTE. Bold variables were used in the multi-group confirmatory factor analysis

To estimate factor scores to be used in the moderation analysis, we estimated factor loadings with a confirmatory multigroup factor analysis, constraining factor loadings and intercepts to be identical across panels. Although, comparing models empirically to determine measurement invariance would suggest, that our data fulfill weak, but not strong measurement invariance. Given that we assume that the different panels are parts of the same general population, albeit of extremely different regions, we expected them to use the measures in the same way. In the following we will use the factor scores from the grouped factor analysis assuming measurement invariance.

2.4 Regression Tables for Study 1

Table S10 – Default Effect

Regression coefficients and standard errors predicting organ donor choice in the default paradigm of the regression models with panel effects (1), only FACE factor variables (2), or panel effects and FACE factor variables.

	Only Panels (1)	Only FACE (2)	FACE & Panels (3)
	B(SE)	B(SE)	B(SE)
Cond (neutral)	.222 (.211)	.770*** (.076)	.744*** (.278)
Cond (treat)	.398* (.213)	1.055*** (.077)	.796*** (.278)
F		.432*** (.067)	.456*** (.090)
A		.310*** (.112)	.274** (.117)
C		.020 (.045)	.027 (.050)
E		-.229*** (.070)	-.178** (.077)
Cond (neutral):F		-.187* (.096)	-.267** (.132)
Cond (treat):F		-.180* (.100)	-.084 (.136)
Cond (neutral):A		-.393** (.164)	-.357** (.172)
Cond (treat):A		-.739*** (.167)	-.670*** (.175)
Cond (neutral):C		.139** (.067)	.099 (.073)
Cond (treat):C		.118* (.069)	.117 (.078)
Cond (neutral):E		.432*** (.101)	.365*** (.110)
Cond (treat):E		.262** (.102)	.254** (.114)
Constant	.486*** (.150)	-.145*** (.052)	-.122 (.193)
Observations	6,438	6,438	6,438
Log Likelihood	-3,987.16	-3,964.44	-3,929.28
Akaike Inf. Crit.	8,040.32	7,958.88	7,948.56

Note: Regression coefficients of panel effects and interactions in (2) and (3) are not depicted for readability. Treat (1) = Neutral Condition, Treat (2) = Opt-out condition. *p<.1; **p<.05; ***p<.01

Table S11 – Framing Effect

Regression coefficients and standard errors predicting risky choice in the framing paradigm of the regression models with panel effects (1), only FACE factor variables (2), or panel effects and FACE factor variables.

	Only Panels (1)	Only FACE (2)	FACE & Panels (3)
	B(SE)	B(SE)	B(SE)
Cond (treat)	1.114*** (.169)	1.196*** (.064)	1.479*** (.227)
F		.076 (.057)	.321*** (.081)
A		.494*** (.103)	.482*** (.107)
C		-.241*** (.042)	-.250*** (.047)
E		-.007 (.060)	-.058 (.067)
Cond (treat):F		.001 (.078)	-.216** (.109)
Cond (treat):A		.087 (.138)	.038 (.144)
Cond (treat):C		.150*** (.057)	.168*** (.063)
Cond (treat):E		-.362*** (.083)	-.200** (.091)
Constant	-.504*** (.119)	-.878*** (.047)	-1.211*** (.164)
Observations	6,438	6,438	6,438
Log Likelihood	-4,043.04	-3,995.83	-3,955.52
Akaike Inf. Crit.	8,130.08	8,011.67	7,971.03

Note: Regression coefficients of panel effects and interactions in (2) and (3) are not depicted for readability. Treat (1) = Loss condition. *p<.1; **p<.05; ***p<.01

Table S12 – Less is better

Regression coefficients and standard errors predicting perceived generosity in the less-is-better paradigm of the regression models with panel effects (1), only FACE factor variables (2), or panel effects and FACE factor variables.

	Only Panels (1)	Only FACE (2)	FACE & Panels (3)
	B(SE)	B(SE)	B(SE)
Cond (treat)	.766*** (.106)	1.016*** (.038)	.412*** (.134)
F		.083** (.034)	-.048 (.045)
A		.435*** (.055)	.349*** (.057)
C		-.056** (.023)	.009 (.025)
E		-.033 (.034)	-.011 (.037)
Cond (treat):F		.073 (.047)	.196*** (.063)
Cond (treat):A		.370*** (.080)	.438*** (.082)
Cond (treat):C		.069** (.033)	.001 (.036)
Cond (treat):E		-.186*** (.048)	-.134** (.053)
Constant	5.980*** (.075)	5.366*** (.027)	5.962*** (.095)
Observations	6,438	6,438	6,438
R^2	0.189	0.209	0.234
Adjusted R^2	0.187	0.208	0.231

Note: Regression coefficients of panel effects and interactions in (2) and (3) are not depicted for readability. Treat (1) = Scarf condition, with the less expensive gift. *p<.1; **p<.05; ***p<.01

Table S13 - Sunk Cost Effect

Regression coefficients and standard errors predicting self-reported likelihood to go the match in the sunk cost paradigm of the regression models with panel effects (1), only FACE factor variables (2), or panel effects and FACE factor variables.

	Only Panels (1)	Only FACE (2)	FACE & Panels (3)
	B(SE)	B(SE)	B(SE)
Cond (treat)	.242 (.222)	.558*** (.083)	.176 (.286)
F		.795*** (.071)	.376*** (.095)
A		.628*** (.121)	.535*** (.124)
C		-.113** (.051)	.069 (.055)
E		.341*** (.073)	.176** (.079)
Cond (treat):F		.073 (.101)	.042 (.135)

Cond (treat):A		-.009 (.172)	.017 (.175)
Cond (treat):C		.029 (.071)	.002 (.076)
Cond (treat):E		-.236** (.104)	-.058 (.112)
Constant	8.159*** (.157)	5.757*** (.059)	7.674*** (.203)
Observations	6,438	6,438	6,438
R^2	0.099	0.063	0.111
Adjusted R^2	0.096	0.061	0.107

Note: Regression coefficients of panel effects and interactions in (2) and (3) are not depicted for readability. Treat (1) = paid condition. *p<.1; **p<.05; ***p<.01

2.5 Regression Coefficients Seemingly Unrelated Regressions

Table S14
Regression coefficients of the seemingly unrelated regression.

	Study 1	
	Sunk Cost	Less-is-better
	β	β
Constant	7.674***	5.962***
Treat	.176	.412**
F	.376***	-.048
A	.535***	.349***
C	.069	.009
E	.176*	-.011
F: Cond (treat)	.042	.196**
A: Cond (treat)	.017	.438***
C: Cond (treat)	.002	.001
E: Cond (treat)	-.058	-.134*
R^2	.11	.23
	Study 4	
Constant	6.12 ***	5.48***
Treat	1.03 ***	1.25***
F	0.43	0.11
A	-0.05	0.06
C	-3.49 .	0.55
E	-0.29	-0.25*
F: Cond (treat)	0.47	0.26
A: Cond (treat)	0.54	0.37*
C: Cond (treat)	1.63	0.98
E: Cond (treat)	-0.10	-0.07
R^2		

Note. Seemingly related regression models can only consist of linear regression models, thus the framing and default effects are not included in the regression system. Panel main effects and interactions are omitted from the Table for readability. The asterisks indicate the *p*-value. * $p < .05$, ** $p < .01$, *** $p < .05$

2.6. Path Coefficients in the Structural Equation Model

Table S15
Path Coefficients of the Structural Equation Models on the Outcomes

	Default	Framing	Less is Better	Sunk Cost
	B	B	B	B
Cond (treat)	.326***	.739***	1.018***	.559***
F	.243	-.061	-.038	.395
A	.482*	-.117	-.297*	.675
C	-.098	-.100	-.071	-.193
E	-.290*	.142	.108	.540
F: Cond (treat)	-.059*	-.005	.071	.078
A: Cond (treat)	-.231***	.058	.372***	-.006
C: Cond (treat)	.034	.087*	.068	.030
E: Cond (treat)	.083**	-.227***	-.183**	-.231*
		Study 4		
Cond (treat)	.141***	.329***	1.268***	1.025***
F	-.224	-.387*	-1.29*	-3.169**
A	-.177	.131	-.401	.188
C	1.478	.67	6.099	5.207
E	-.328**	.004	-.2	-.787
F: Cond (treat)	.102*	-.031	.168	.156
A: Cond (treat)	-.021	.005	.382*	.632*
C: Cond (treat)	-1.03***	.347	1.234	2.678
E: Cond (treat)	.036	-.179***	-.097	-.318

Note. Choices in the default and framing paradigms were modelled as ordinal variables, thresholds are not reported in this panel for visibility. Panel effects were not included in the model. The asterisks indicate the *p*-value. * $p < .05$, ** $p < .01$, *** $p < .05$

3 Additional Material for Study 3

3.1 Experimental Design. Study 3 was a preregistered extension of Study 1 (doi.org/1.17605/OSF.IO/RC963) that included new behavioral treatments and measures of economic preferences. Additionally, we varied sampling periods systematically across different days of the week and times of day in order to investigate the possibility that distributions of our proposed FACE factor variables differ over the course of the day and week. The trolley dilemma, replicated by Many Labs 2 (Klein et al. 2018), asks individuals to assess the moral permissibility of inaction (with severe consequences) verses action (with lessened consequences). We use two distinct scenarios, one involving the decision of a passenger on a train with broken brakes to do nothing (killing the five people on the tracks ahead) or to turn the train (killing the one person on the other track); and another involving the decision of a bystander to sacrifice the life of one man in order to save five others, or spare the individual and allow the five others to be killed. This study also included the false consensus effect, quantified by meta-analysis (Mullen et al. 1984),

describes individuals' tendency to overestimate the degree of consensus regarding their decisions in a hypothetical scenario (Ross et al., 1977). To measure economic preferences, we adopt risk preference measures, time preference measures, and reciprocity measures from Falk et al. (2018). The analysis of these measures is not included in this project. Economic preferences and the false consensus paradigm were not further analyzed in the manuscript to focus on experimental paradigms only.

In order to increase the likelihood of variation of the distribution of FACE factor variables, we define six sampling periods (at which we collected data from both panels). On both Sunday and Wednesday, we initiated survey sampling at 1:00 am PST (4:00 am EST), 8:00 am PST (12:00 pm EST), and 5:00 pm PST (8:00 pm EST). We selected these times such that each period captured respondents in the late-night hours, during early working hours, and during evening hours, respectively, regardless of where in the United States they were located. We collected N = 100 respondents from each panel, at each time (on each day) for a total of N = 1200 prior to filtering. Finally, we augment the cognitive and numeric measures of Study 1 and 2 by including Raven's-like matrix measures and 3D rotation tasks (Condon & Revelle, 2014).

Figure S1. Average FACE factor scores per panel and time of the day in Study 3.

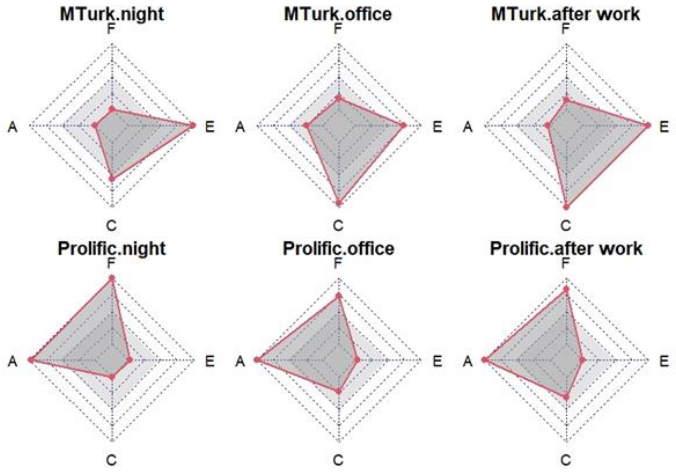


Table S16

Effect sizes (η^2) for the main effects and interactions of condition (cond), day of the week (day), time of the day (timeofday) and panel effects in Study 3, (N = 1,196)

Effect	Cond	Panel	Panel: Cond	Day	Day: Cond	TimeofDay	TimeofDay: Cond
Trolley Problem	**.134*****	**.058*****	**.031*****	.003*	.000	**.004***	.002
	(.138*****)**	**(.003*****)**	**(.010*******)**	(.004*)	(.000)	(.004)	(.002)

NOTE. The Table reports the effects sizes in the complete data in the absence (without FACE factor variables) and presence of FACE factors in parentheses and their interactions with the treatment effects in the models. Significant effects are bold and the asterisks indicate the *p*-value.
* $p < .05$, ** $p < .01$, *** $p < .05$

3.2 Regression Table for Study 3

Table S17
Trolley paradigm. Regression coefficients and standard errors predicting the probability to act of the regression models with panel and temporal effects (1), only FACE factor variables (2), or panel and temporal effects and FACE factor variables.

	Only Panels (1)	Only FACE (2)	FACE & Panels (3)
	B(SE)	B(SE)	B(SE)
Cond (treat)	.815** (.346)	1.240*** (.206)	.767** (.372)
F		-.375* (.223)	-.311 (.239)
A		-1.617*** (.234)	-1.369*** (.241)
C		.285*** (.102)	.253** (.108)
E		.668*** (.163)	.389** (.177)
Cond (treat):F		.465 (.336)	.324 (.355)
Cond (treat):A		1.423*** (.337)	1.081*** (.348)
Cond (treat):C		-.140 (.150)	-.049 (.157)
Cond (treat):E		-.298 (.253)	.036 (.261)
Constant	1.261*** (.217)	.680*** (.134)	1.441*** (.245)
Observations	1,196	1,196	1,196
Log Likelihood	-598.893	-573.694	-557.122
Akaike Inf. Crit.	1,217.79	1,167.39	1,150.24

N*ote:* Regression coefficients of panel day and time of the week effects and interactions in (2) and (3) are not depicted for readability. Treat (1) = tracks condition. *p<.1; **p<.05; ***p<.01

4 Additional Material for Study 4

4.1 Experimental Design. In Study 4 (preregistered at https://osf.io/s5wju), we adapted the questionnaires used in the Study 1 series by dropping the local warming paradigm and incorporating the 10-item synonym task, adapted from CREATE's Common Core Battery of Measures (Czaja et al. 2006; Czaja et al. 2006; Li et al. 2013) and a 10-item antonym task (Salthouse, 1993) as additional measures for crystallized intelligence.

4.2 Sampling Details. As pre-registered, in Study 4, we aimed to collect 500 complete responses from each of three platforms: MTurk, Lucid Marketplace (blended), and Prolific, totaling 1500 complete responses. These were to be gathered via a Qualtrics survey deployed in four separate sampling periods (waves). The compensation was structured as follows: $2 for MTurk respondents, $2.5 for Lucid respondents, and $3 for Prolific respondents (also see Table S1). Due to the additional synonym and antonym tasks and thus the anticipated longer completion time relative to Study 1 series, compensation rates were proportionally adjusted across panels to maintain respondent motivation and ensure fair compensation.* The survey was configured to target a standard population within each panel.† Prospective respondents saw the survey described as a “general population survey,” accompanied by the introductory note: “Welcome! We appreciate your willingness to participate in this brief survey. This survey contains various

* Only respondents who fully completed the survey were eligible for compensation.
† For Prolific, the survey details specified that it was intended for a “standard sample.” On MTurk, the only qualification for participation was the exclusion of individuals who had participated in our prior studies; we did not request "master workers" or specify any HIT approval rate. On Lucid, it was stipulated that only adults were eligible to participate.

unrelated questions from different tasks. Please answer each question independently of the others." Respondents were informed that the survey would require approximately 15 minutes to complete.

We varied sampling periods systematically across different days of the week and times of day to increase the likelihood of variation of the distribution of FACE factor variables. We initiated survey sampling at 8:30 am PST (11:30 am EST) on July 26, 9:00 am PST (12:00 pm EST) on July 28, 8:00 am PST (11:00 am EST) on July 30, and 8:30 am PST (11:30 am EST) on July 31. On a given data collection day, data collection started in each panel simultaneously, within a 60-minute window to allow for execution.

In Study 4, we configured all three panels to ensure that no respondent should be part of more than one wave. An ensuing analysis, using unique identifiers from each panel, revealed an absence of duplicate respondents. These findings indicate that our exclusion criteria were effectively implemented, ensuring no respondent took our survey more than once across the three panels. The procedural setup remained consistent across all four waves.

Conducting the study over four days allowed us to diversify our pool of respondents within each panel[‡]. In each wave, we aimed to collect 125 complete responses, totaling 500 complete responses per panel. Our final sample size before filtering was 480 in MTurk, 505 in Prolific, and 500 in Lucid Marketplace (also see Table S1). Deviations from our target of 500 complete responses per panel per wave were due to factors such as respondent dropouts[§] or incomplete verification processes with the panel providers[**]. These were factors that were beyond our control.

Figure S2. Average FACE factor scores per panel in Study 4.

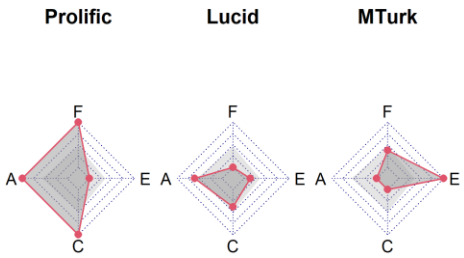
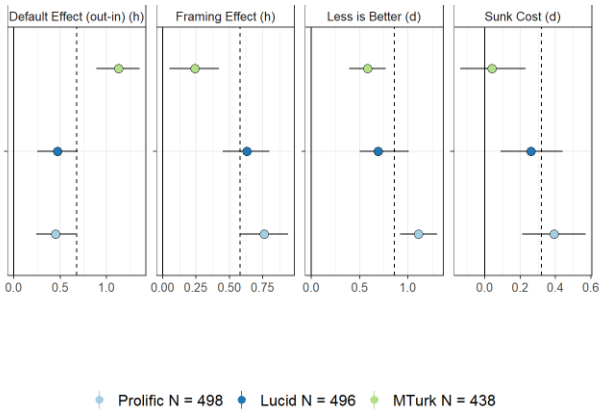


Figure S3. Average Effect Sizes per Panel in Study 4.

[‡] This approach was particularly relevant for Lucid, which unlike MTurk and Prolific, operates as a panel broker that channels surveys to various commercial panels like Tap Research and Qmee. Our data indicate that the influx from each panel supplier is typically homogeneous within specific time windows. By spreading data collection over four days, we increased the diversity of panel suppliers. For detailed breakdowns, refer to Figures 1)

[§] Dropout rates across the panels were observed as follows: 10% for MTurk, 12.4% for Lucid, and 3% for Prolific. For this study, "dropouts" were defined as respondents who provided informed consent, completed the first attention check question, but did not complete it in entirety. A potential contributing factor for these dropout rates could be the perceived challenge associated with the synonym and antonym tasks.

[**] For example, some participants failed to provide the required verification codes to the panel suppliers.



4.3 Regression Tables for Study 4

Table S18 – Default Effect

Regression coefficients and standard errors predicting organ donor choice in the default paradigm of the regression models with panel effects (1), panel effects and demographic variables (2), panels effects and FACE factor variables (3), or panels effects and TIPI variables (4).

	Only Panels (1)	Panels & Demographics (2)	Panels & FACE (3)	Panels & TIPI (4)
	B(SE)	B(SE)	B(SE)	B(SE)
Cond(neutral)	.726*** (.232)	1.071*** (.302)	.600** (.240)	2.485*** (.964)
Cond(treat)	.965*** (.240)	1.162*** (.299)	.861*** (.246)	.817 (.960)
Edu		.245** (.109)		
Gender		.514** (.202)		
Age		.023 (.104)		
F			-.438 (.408)	
A			.176 (.235)	
C			7.302*** (1.989)	
E			.113 (.202)	
TIPI(E)				.003 (.076)
TIPI(A)				.134 (.101)
TIPI(C)				-.142 (.091)
TIPI(ES)				-.086 (.085)
TIPI(O)				.159* (.090)
Edu:Cond(neutral)		.085 (.164)		
Edu:Cond(treat)		-.103 (.159)		
Cond(neutral):Gender		-.533* (.304)		
Cond(treat):Gender		-.334 (.301)		
Cond(neutral):Age		.084 (.160)		
Cond(treat):Age		-.134 (.153)		
F:Cond(neutral)			-.540 (.671)	
F:Cond(treat)			.787 (.652)	
Cond(neutral):A			-.651* (.377)	
Cond(treat):A			-.315 (.368)	
Cond(neutral):C			-1.406 (2.999)	
Cond(treat):C			-6.570** (2.978)	
Cond(neutral):E			-.471 (.336)	

Cond(treat):E			-.143 (.319)	
TIPI(E):Cond(neutral)				.199* (.113)
TIPI(E):Cond(treat)				.132 (.118)
Cond(neutral):TIPI(A)				.089 (.162)
Cond(treat):TIPI(A)				.157 (.152)
Cond(neutral):TIPI(C)				-.165 (.146)
Cond(treat):TIPI(C)				-.012 (.147)
Cond(neutral):TIPI(ES)				-.060 (.127)
Cond(treat):TIPI(ES)				-.129 (.132)
Cond(neutral):TIPI(O)				-.331** (.138)
Cond(treat):TIPI(O)				-.068 (.139)
Constant	.182 (.156)	-.114 (.206)	.291* (.164)	-.222 (.641)
Observations	1,432	1,432	1,432	1,432
Log Likelihood	-821.883	-809.928	-805.144	-800.518
Akaike Inf. Crit.	1,661.77	1,655.86	1,652.29	1,649.04

Note: Regression coefficients of panel effects and interactions in (1) and (2) are not depicted for readability. Treat (1) = Neutral Condition, Treat (2) = Opt-out condition. TIPI (E) = Extraversion, TIPI (A) = Agreeableness, TIPI (C) = Conscientiousness, TIPI (ES) = Emotional Stability, TIPI (O) = Openness. *p<.1; **p<.05; ***p<.01

Table S19 – Framing Effect

Regression coefficients and standard errors predicting risky choice in the framing paradigm of the regression models with panel effects (1), panel effects and demographic variables (2), panels effects and FACE factor variables (3), or panels effects and TIPI variables (4).

	Only Panels (1)	Panels & Demographics (2)	Panels & FACE (3)	Panels & TIPI (4)
	B(SE)	B(SE)	B(SE)	B(SE)
Cond(treat)	1.534***(.244)	1.641*** (.202)	2.084*** (.782)	2.508*** (.833)
Edu	-.048 (.094)			
Gender	-.033 (.176)			
Age	.002 (.092)			
F		.784* (.403)		.847** (.411)
A		.747*** (.219)		.575** (.225)
C		-3.729** (1.807)		-4.097** (1.834)
E		.121 (.195)		.249 (.198)
TIPI(E)			.042 (.070)	.066 (.071)
TIPI(A)			.033 (.094)	-.010 (.097)
TIPI(C)			.173** (.088)	.129 (.091)
TIPI(ES)			.157** (.078)	.141* (.078)
TIPI(O)			.120 (.080)	.110 (.081)
Edu:Cond(treat)		-.084 (.126)		
Cond(treat):Gender		.150 (.238)		
Cond(treat):Age		.137 (.124)		
F:Cond(treat)			-.272 (.535)	
Cond(treat):A			-.048 (.301)	
Cond(treat):C			1.487 (2.404)	
Cond(treat):E			-.685*** (.263)	
TIPI(E):Cond(treat)				-.122 (.093)
Cond(treat):TIPI(A)				.132 (.127)
Cond(treat):TIPI(C)				-.058 (.118)
Cond(treat):TIPI(ES)				-.129 (.105)
Cond(treat):TIPI(O)				.031 (.114)

Constant	-1.222***(.152)	-1.204***(.183)	-1.260*** (.155)	-3.756*** (.599)
Observations	1,432	1,432	1,432	1,432
Log Likelihood	-880.713	-877.721	-839.574	-855.351
Akaike Inf. Crit.	1,773.43	1,779.44	1,707.15	1,742.70

Note: Regression coefficients of panel effects and interactions in (1)-(4) are not depicted for readability. Treat (1) = Loss condition. TIPI (E) = Extraversion, TIPI (A) = Agreeableness, TIPI (C) = Conscientiousness, TIPI (ES) = Emotional Stability, TIPI (O) = Openness. *p<.1; **p<.05; ***p<.01

Table S20 – Less is Better

Regression coefficients and standard errors predicting perceived generosity in the less-is-better paradigm of the regression models with panel effects (1), panel effects and demographic variables (2), panels effects and FACE factor variables (3), or panels effects and TIPI variables (4).

	Only Panels (1)	Panels & Demographics (2)	Panels & FACE (3)	Panels & TIPI (4)
	B(SE)	B(SE)	B(SE)	B(SE)
Cond(treat)	1.272***(.123)	1.140*** (.152)	1.255*** (.119)	-.243 (.469)
Edu		-.154*** (.053)		
Gender		.108 (.105)		
Age		.066 (.055)		
F			.114 (.212)	
A			.057 (.121)	
C			.547 (1.021)	
E			-.249** (.101)	
TIPI(E)				-.064 (.041)
TIPI(A)				.160*** (.058)
TIPI(C)				-.118** (.050)
TIPI(ES)				.057 (.045)
TIPI(O)				-.021 (.051)
Edu:Cond(treat)		.137* (.079)		
Cond(treat):Gender		.263* (.149)		
Cond(treat):Age		.050 (.078)		
F:Cond(treat)			.261 (.303)	
Cond(treat):A			.373** (.172)	
Cond(treat):C			.975 (1.407)	
Cond(treat):E			-.072 (.149)	
TIPI(E):Cond(treat)				.008 (.057)
Cond(treat):TIPI(A)				.006 (.078)
Cond(treat):TIPI(C)				.237*** (.071)
Cond(treat):TIPI(ES)				-.108* (.064)
Cond(treat):TIPI(O)				.147** (.069)
Constant	5.476***(.087)	5.423*** (.110)	5.479*** (.084)	5.320*** (.333)
Observations	1,432	1,432	1,432	1,432
R^2	0.202	0.218	0.264	0.234
Adjusted R^2	0.199	0.212	0.258	0.226

Note: Regression coefficients of panel effects and interactions in (1)-(4) are not depicted for readability. Treat (1) = Scarf condition, with the less expensive gift. TIPI (E) = Extraversion, TIPI (A) = Agreeableness, TIPI (C) = Conscientiousness, TIPI (ES) = Emotional Stability, TIPI (O) = Openness. *p<.1; **p<.05; ***p<.01

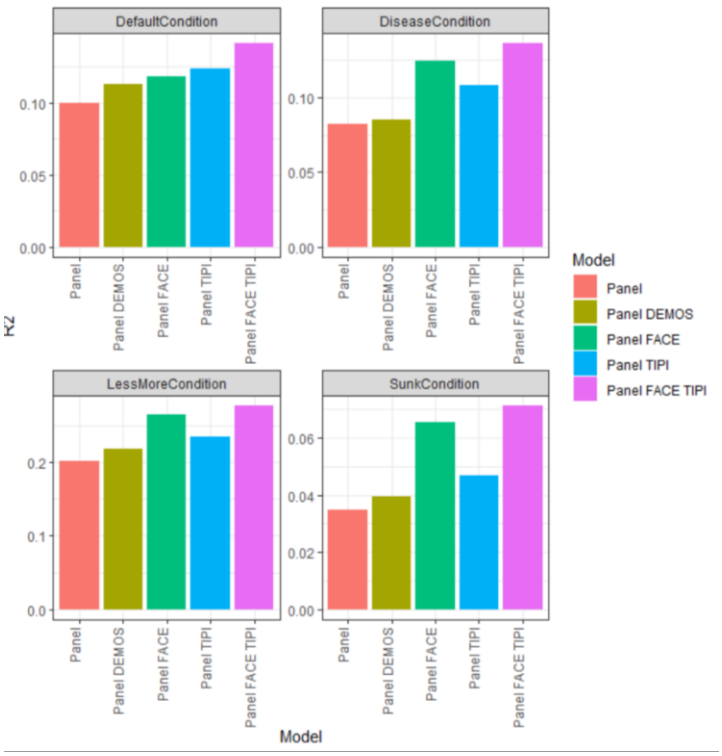
Table S21 – Sunk Cost Effect

Regression coefficients and standard errors predicting self-reported likelihood to go the match in the sunk cost paradigm of the regression models with panel effects (1), panel effects and demographic variables (2), panels effects and FACE factor variables (3), or panels effects and TIPI variables (4).

	Only Panels (1)	Panels & Demographics (2)	Panels & FACE (3)	Panels & TIPI (4)
	B(SE)	B(SE)	B(SE)	B(SE)
Cond(treat)	1.005*** (.233)	.899*** (.289)	1.027*** (.230)	-.337 (.899)
Edu		.087 (.102)		
Gender		-.025 (.200)		
Age		-.135 (.103)		
F			.431 (.414)	
A			-.047 (.233)	
C			-3.486* (1.895)	
E			-.288 (.203)	
TIPI(E)				.074 (.076)
TIPI(A)				.083 (.107)
TIPI(C)				-.058 (.100)
TIPI(ES)				-.029 (.087)
TIPI(O)				-.020 (.089)
Edu:Cond(treat)		-.262* (.151)		
Cond(treat):Gender		.263 (.283)		
Cond(treat):Age		.223 (.149)		
F:Cond(treat)				
Cond(treat):A			.471 (.586)	
Cond(treat):C			.545 (.334)	
Cond(treat):E			1.626 (2.727)	
TIPI(E):Cond(treat)			-.105 (.288)	
Cond(treat):TIPI(A)				-.116 (.110)
Cond(treat):TIPI(C)				.270* (.150)
Cond(treat):TIPI(ES)				.128 (.137)
Cond(treat):TIPI(O)				-.080 (.123)
Constant				.011 (.133)
Observations	1,432	1,432	1,432	1,432
R^2	0.035	0.039	0.065	0.047
Adjusted R^2	0.031	0.032	0.057	0.037

Note: Regression coefficients of panel effects and interactions in (1)-(4) are not depicted for readability. Treat (1) = paid condition. TIPI (E) = Extraversion, TIPI (A) = Agreeableness, TIPI (C) = Conscientiousness, TIPI (ES) = Emotional Stability, TIPI (O) = Openness. *p<.1; **p<.05; ***p<.01

Figure S4. Explained Variance in the Outcome Variables Study 4.



5 Additional Material for Study 5

5.1 Experimental Design. Study 5 was a preregistered (https://osf.io/gxcyd) extension of Study 4 aiming to examine the effect of moderators on the manipulation intensity, and to enhance the fluid intelligence measures. Thus, we adapted the questionnaire used in Study 4 by including measures of the received condition, dropping the Numeracy Task, and including the Raven's-like matrix measures and 3D rotation tasks (Condon & Revelle, 2014) used in Study 3.

Table S22 – Received Condition

Exact wording of the received condition measures administered after all paradigms had been presented.

Paradigm	Received Condition Measure	
Default	You were asked to imagine that you needed to get a new driver's license and had to make a decision about being an organ donor. Which of the following (A or B) best describes the question you saw?	
	A: *In your state, every person was considered not to be an organ donor unless they choose to be. You could accept this default or opt into being an organ donor.*	B**:** *In your state, every person was considered to be an organ donor unless they choose not to be. You could accept this default or opt out of being an organ donor.*
Less is better	You were asked to imagine that you received a goodbye gift from a friend. Which of the following (A or B) best describes the question you saw?	
	A: *You were given a wool coat from a department store that sells coats between $100 and $1000, and this one was $110.*	**B:** *You were given a wool scarf from a department store that sells scarves between $10 and $100, and this one was $90.*
Sunk Cost	You were asked to imagine that you had tickets to see your favorite football team playing an important game, but that it was freezing cold on game day. Which of the following (A or B) best describes the question you saw?	
	A: *You paid handsomely for your ticket.*	**B:** *You received a ticket for free from a friend.*

Framing (unusual disease)	You were asked to imagine that the United States was preparing for the outbreak of an unusual disease. Which of the following (A or B) best describes the question you saw?	
	A: *You chose between two programs. In one program, 200 people would be saved. In the other, there was 1/3 probability that 600 people would be saved and 2/3 probability that no people would be saved.*	**B:** *You chose between two programs. In one program, 400 people would die. In the other, there was 1/3 probability that no people would die and 2/3 probability that 600 people would die.*
Note: **Respondents answered on a 5-point Likert scale from 1 (I definitely saw A) to 5 (I definitely saw B).**		

5.2 Sampling Details. As pre-registered, in Study 5, we aimed to collect 500 complete responses from each of three platforms: MTurk, Lucid Marketplace (blended), and Prolific, totaling 1500 complete responses. The content of the Qualtrics survey remained consistent across all waves. Respondents saw the same title and description of the survey as those Study 4.

The responses were to be gathered via a Qualtrics survey deployed in four separate sampling periods (waves). We varied sampling periods systematically across different days of the week and times of day to increase the likelihood of variation of the distribution of FACE factor variables. The survey waves were launched over four consecutive days, from August 15, 2023, to August 18, 2023. We initiated survey sampling at 8:30 am PST (11:30 am EST) on August 15, 11:30 am PST (2:30 pm EST) on August 16, 12:00 pm PST (3:00 pm EST) on August 17, and 6:30 am PST (9:30 am EST) on August 18. On a given data collection day, data collection started in each panel simultaneously, within a 60-minute window to allow for execution. In Study 5, we configured all three panels to exclude respondents who had previously taken any surveys in this ongoing online research series, and that no respondent should be part of more than one wave. An ensuing analysis indicate that our exclusion criteria were effectively implemented. The procedural setup remained consistent across all four waves.

Conducting the study over four days allowed us to diversify our pool of respondents within each panel. In each wave, we aimed to collect 125 complete responses, totaling 500 complete responses per panel. Our final sample size before filtering was 510 in MTurk, 504 in Prolific, and 496 in Lucid Marketplace (also see Table S1). Deviations from our target of 500 complete responses per panel per wave were due to factors such as respondent dropouts†† or incomplete verification processes with the panel providers‡‡. These were factors that were beyond our control.

Figure S5. Average FACE factor scores per panel in Study 5.

†† Dropout rates across the panels were observed as follows: 10% for MTurk, 12.4% for Lucid, and 3% for Prolific. For this study, “dropouts” were defined as respondents who provided informed consent, completed the first attention check question, but did not complete it in entirety. A potential contributing factor for these dropout rates could be the perceived challenge associated with the synonym and antonym tasks.

‡‡ For example, some participants failed to provide the required verification codes to the panel suppliers.

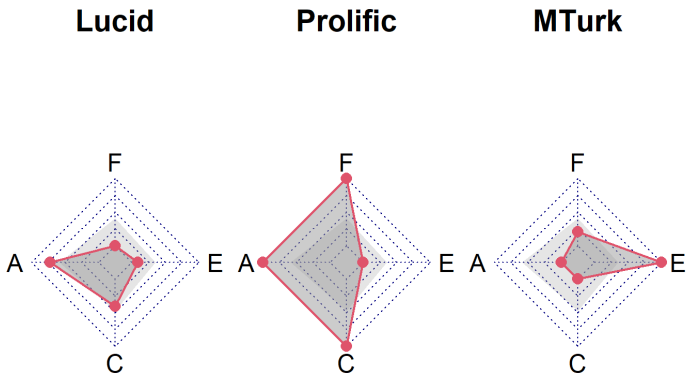


Figure S6. Average Effect Sizes per Panel in Study 5.

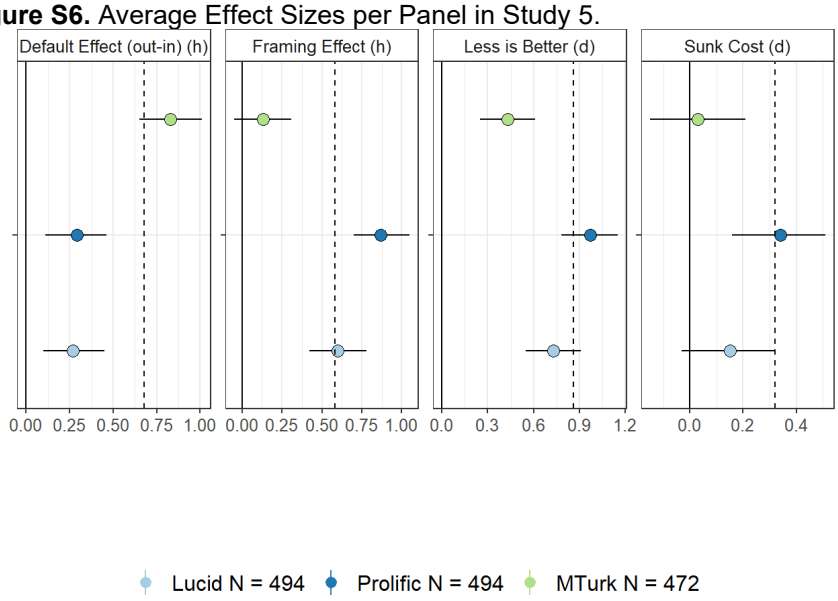
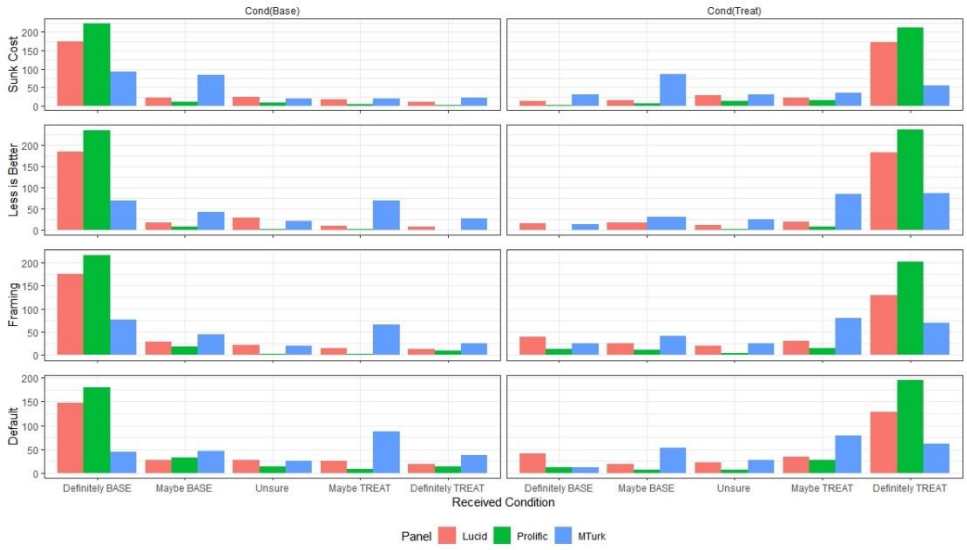


Figure S7. Received Condition per Panel and Paradigm in Study 5.



5.4 Moderation of Manipulation Intensity: Received Condition Regression Tables for Study 5

Table S23 – Linear Regression Models

Regression coefficients and standard errors of linear regression models predicting the received condition in all four paradigms in Study 5 using the assigned condition and FACE moderators.

	Received Condition			
	Default	Less is better	Sunk cost	Framing
	B(SE)	B(SE)	B(SE)	B(SE)
Cond (Treat)	-.072** (.031)	-.062** (.028)	-.075** (.038)	.003 (.040)
F	1.496*** (.048)	1.497*** (.043)	1.156*** (.058)	1.052*** (.061)
A	-.103*** (.024)	-.220*** (.023)	-.139*** (.030)	-.116*** (.030)
C	-.095*** (.035)	-.111*** (.033)	-.082* (.043)	-.205*** (.045)
E	.079*** (.018)	.136*** (.017)	.130*** (.022)	.122*** (.022)
Cond(treat):F	.059 (.044)	.154*** (.040)	.217*** (.053)	.159*** (.056)
Cond(treat):A	.301*** (.035)	.319*** (.031)	.143*** (.042)	.121*** (.045)
Cond(treat):C	.290*** (.050)	.230*** (.045)	.152** (.060)	.320*** (.063)
Cond(treat):E	-.225*** (.026)	-.198*** (.023)	-.170*** (.031)	-.185*** (.033)
Constant	-.686*** (.034)	-.788*** (.030)	-.617*** (.041)	-.606*** (.043)
R^2	.644	.714	.478	.422

Note: Received Conditions were coded as continuous variables, higher values indicating self-reported reception of the Treatment Condition. *p<.1; **p<.05; ***p<.01

Table S24 – Multinomial Logistic Regression Model: Default Paradigm

Regression coefficients and standard errors of multinomial logistic regression models predicting the received condition in the default paradigm using the assigned condition, and FACE moderators.

	Received Condition Response			
	Cond(base)[1]	Unsure	Cond(treat)[1]	Cond(treat)[2]
	B(SE)	B(SE)	B(SE)	B(SE)
Cond(treat)	.608* (.315)	1.113*** (.338)	1.707*** (.289)	2.999*** (.277)
F	.032 (.160)	-.173 (.194)	.064 (.167)	.037 (.193)
A	-.240* (.126)	-.212 (.144)	-.425*** (.124)	-.364** (.146)
C	-.323* (.184)	-.968*** (.274)	-.974*** (.230)	-.443* (.231)
E	.374*** (.092)	.228** (.107)	.476*** (.089)	.389*** (.105)
F:Cond(treat)	.907*** (.300)	.792** (.331)	.689** (.281)	.948*** (.281)
A:Cond(treat)	.050 (.223)	.155 (.245)	.222 (.208)	.357 (.219)
C:Cond(treat)	-.807** (.374)	-.027 (.442)	.114 (.360)	.352 (.327)
E:Cond(treat)	-.086 (.156)	-.087 (.173)	-.213 (.145)	-.588*** (.159)
Constant	-1.332*** (.177)	-1.791*** (.216)	-1.643*** (.201)	-1.842*** (.218)
Note: *p<.1; **p<.05; ***p<.01				

Table S25 – Multinomial Logistic Regression Model: Framing Paradigm

Regression coefficients and standard errors of multinomial logistic regression models predicting the received condition in the framing paradigm using the assigned condition, and FACE moderators.

	Received Condition Response			
	Cond(base)[1]	Unsure	Cond(treat)[1]	Cond(treat)[2]
Cond(treat)	1.161*** (.312)	1.096*** (.356)	2.434*** (.343)	3.375*** (.283)
F	-.152 (.176)	-.744*** (.248)	.097 (.203)	-.480** (.231)
A	-.455*** (.134)	-.263 (.178)	-.715*** (.146)	-.384** (.170)
C	-.572** (.225)	-.112 (.304)	-1.031*** (.295)	-.023 (.279)
E	.390*** (.092)	.287** (.120)	.554*** (.100)	.417*** (.113)
F:Cond(treat)	.589** (.291)	1.031*** (.361)	.405 (.291)	1.095*** (.291)
A:Cond(treat)	-.066 (.221)	-.242 (.265)	.179 (.216)	.211 (.222)
C:Cond(treat)	.098 (.368)	-.519 (.465)	.582 (.390)	.040 (.346)
E:Cond(treat)	-.199 (.149)	-.201 (.178)	-.366** (.147)	-.505*** (.151)
Constant	-1.749*** (.195)	-2.060*** (.221)	-2.560*** (.269)	-2.142*** (.229)
Akaike Inf. Crit.	3,164.53			
Note: *p<.1; **p<.05; ***p<.01				

Table S26 – Multinomial Logistic Regression Model: Less is better Paradigm

Regression coefficients and standard errors of multinomial logistic regression models predicting the received condition in the Less is better using the assigned condition, and FACE moderators.

	Received Condition Response			
	Cond(base)[1]	Unsure	Cond(treat)[1]	Cond(treat)[2]
	B(SE)	B(SE)	B(SE)	B(SE)
Cond(treat)	2.688*** (.534)	2.211*** (.544)	3.911*** (.524)	6.421*** (.550)
F	-.082 (.212)	-.600** (.234)	.022 (.233)	-.332 (.281)
A	-.741*** (.165)	-.724*** (.178)	-1.222*** (.178)	-.935*** (.207)
C	-1.019*** (.320)	-.841** (.355)	-.916*** (.355)	-1.404*** (.499)

E	.693*** (.107)	.392*** (.120)	.866*** (.116)	.532*** (.138)
F:Cond(treat)	.865** (.441)	1.624*** (.462)	1.187*** (.423)	1.369*** (.438)
A:Cond(treat)	.506* (.281)	.209 (.304)	.945*** (.271)	1.083*** (.286)
C:Cond(treat)	.760 (.698)	.737 (.720)	.729 (.652)	2.414*** (.716)
E:Cond(treat)	-.552*** (.188)	-.338* (.205)	-.611*** (.182)	-.737*** (.192)
Constant	-2.550*** (.274)	-2.325*** (.265)	-3.131*** (.331)	-3.398*** (.416)
Akaike Inf. Crit.	2,325.32			

Note: *p<.1; **p<.05; ***p<.01

Table S27 – Multinomial Logistic Regression Model: Sunk Cost Paradigm

Regression coefficients and standard errors of multinomial logistic regression models predicting the received condition in the sunk cost paradigm using the assigned condition, and FACE moderators.

	Received Condition			
	Cond(base)[1]	Unsure	Cond(treat)[1]	Cond(treat)[2]
	B(SE)	B(SE)	B(SE)	B(SE)
Cond(treat)	2.180*** (.416)	3.033*** (.411)	3.393*** (.443)	6.065*** (.482)
F	.265* (.161)	-.243 (.211)	-.631** (.271)	-.222 (.280)
A	-.507*** (.120)	-.289* (.154)	-.469*** (.175)	-.350* (.184)
C	-.806*** (.210)	-.718** (.295)	-.319 (.379)	-1.231** (.480)
E	.416*** (.081)	.200* (.106)	.338*** (.118)	.361*** (.125)
F:Cond(treat)	.311 (.320)	.333 (.356)	.852** (.395)	.382 (.376)
A:Cond(treat)	.735*** (.234)	.726*** (.263)	.868*** (.275)	1.297*** (.270)
C:Cond(treat)	.087 (.456)	.525 (.499)	.339 (.545)	1.863*** (.590)
E:Cond(treat)	-.096 (.154)	-.236 (.176)	-.254 (.182)	-.774*** (.182)
Constant	-2.025*** (.203)	-2.283*** (.232)	-2.710*** (.283)	-3.317*** (.383)
Akaike Inf. Crit.	2,770.67			

Note: *p<.1; **p<.05; ***p<.01

5.4 Moderation of the Direct Effect of the Manipulation: Received Condition Regression Tables for Study 5

Table S28 – Regression models to test the FACE factors interactions with the Received Condition on the Outcome Variables.

	Outcome Variable			
	Sunk Cost	Less-is-better	Framing (logistic)	Default (logistic)
	B(SE)	B(SE)	B(SE)	B(SE)
Received Condition	.233** (.102)	.523*** (.054)	.517*** (.089)	-.141* (.082)
F	.468*** (.095)	.015 (.050)	-.109 (.081)	.044 (.076)
A	.107 (.076)	.124*** (.040)	.382*** (.070)	-.113* (.060)
C	-.321*** (.107)	.253*** (.057)	-.035 (.092)	.350*** (.088)
E	-.073 (.056)	-.164*** (.030)	-.269*** (.052)	.021 (.044)

Received Condition:F	.077 (.093)	-.040 (.049)	.061 (.081)	.206*** (.075)
Received Condition:A	.169** (.086)	.068 (.046)	.346*** (.079)	.046 (.067)
Received Condition:C	-.126 (.103)	.067 (.055)	.061 (.090)	-.220** (.086)
Received Condition:E	-.090 (.065)	-.015 (.034)	.046 (.060)	.007 (.051)
Constant	5.923*** (.103)	5.724*** (.055)	-.229*** (.087)	.287*** (.082)

Note: *p<.1; **p<.05; ***p<.01

Table S29 – Regression models to test the direct effects of the manipulation and interaction with FACE factors conditional on the Received Condition

	Sunk Cost	Less-is-better	Framing	Default
	B(SE)	B(SE)	B(SE)	B(SE)
Cond(Treat)	.221 (.267)	.682*** (.147)	.867*** (.194)	.955*** (.188)
F	.374*** (.133)	.095 (.071)	-.082 (.122)	-.120 (.109)
A	.087 (.105)	-.027 (.059)	.409*** (.101)	-.060 (.084)
C	-.300** (.153)	.153* (.082)	-.222 (.139)	.597*** (.127)
E	-.059 (.077)	-.175*** (.043)	-.207*** (.075)	-.023 (.062)
Received Condition	.111 (.114)	.259*** (.066)	.338*** (.079)	-.305*** (.076)
Cond(Treat):F	.176 (.190)	-.170* (.100)	-.035 (.165)	.382** (.158)
Cond(Treat):A	.025 (.155)	.293*** (.082)	.044 (.136)	-.160 (.128)
Cond(Treat):C	-.026 (.218)	.205* (.114)	.345* (.186)	-.508*** (.182)
Cond(Treat):E	-.017 (.114)	.028 (.059)	-.094 (.101)	.155 (.095)
Constant	5.829*** (.166)	5.376*** (.092)	-.707*** (.133)	-.202 (.127)
R^2	.04	.225	.129	.081

Note: *p<.1; **p<.05; ***p<.01

5.5 Regression Tables for Study 5 Comparing Different Moderators
Table S30 – Default Effect

Regression coefficients and standard errors predicting organ donor choice in the default paradigm of the regression models with panel effects (1), panel effects and demographic variables (2), panels effects and FACE factor variables (3), or panels effects and TIPI variables (4).

	Only Panels (1)		Panels & Demographics (2)		Panels & FACE (3)		Panels & TIPI (4)	
	B	SE	B	SE	B	SE	B	SE
Cond (treat)	0.555***	0.184	0.801***	0.258	0.754**	0.299	1.350*	0.737
EDU			0.418***	0.089				
Gender			0.368**	0.158				
Age			0.094	0.085				
F					0.028	0.197		
A					-0.059	0.091		

C					0.534***	0.135		
E					0.05	0.128		
Cond (treat):F					0.245	0.29		
Cond (treat):A					-0.172	0.139		
Cond (treat):C					-0.471**	0.195		
Cond (treat):E					0.131	0.19		
TIPI(E)							-0.094	0.065
TIPI(A)							0.135*	0.08
TIPI(C)							-0.034	0.082
TIPI(ES)							0.033	0.07
TIPI(O)							0.094	0.072
Cond (treat):TIPI(E)							0.187**	0.091
Cond (treat):TIPI(A)							-0.197*	0.12
Cond (treat):TIPI(C)							-0.001	0.117
Cond (treat):TIPI(ES)							-0.124	0.102
Cond (treat):TIPI(O)							0.014	0.107
Constant	0.048	0.127	-0.034	0.18	0.141	0.203	-0.702	0.5
Observations	1,460		1,460		1,460		1,460	
Log Likelihood	-936.11		-918.82		-914.86		-927.68	
Akaike Inf. Crit.	1,884.21		1,861.63		1,857.73		1,887.37	

Note: Regression coefficients of panel effects and interactions in (1) and (2) are not depicted for readability. Treat (1) = Opt-out condition. TIPI (E) = Extraversion, TIPI (A) = Agreeableness, TIPI (C) = Conscientiousness, TIPI (ES) = Emotional Stability, TIPI (O) = Openness. *p<.1; **p<.05; ***p<.01

Table S31 – Framing Effect

Regression coefficients and standard errors predicting risky choice in the framing paradigm of the regression models with panel effects (1), panel effects and demographic variables (2), panels effects and FACE factor variables (3), or panels effects and TIPI variables (4).

	Only Panels (1)		Panels & Demographics (2)		Panels & FACE (3)		Panels & TIPI (4)	
	B	SE	B	SE	B	SE	B	SE
Cond (treat)	1.223***	0.189	1.184***	0.263	1.468***	0.312	-0.324	0.769
EDU			-0.210**	0.095				
Gender			0.045	0.176				
Age			-0.103	0.091				
F					0.084	0.228		
A					0.364***	0.107		

C					-0.318**	0.158		
E					-0.165	0.145		
Cond (treat):F					0.192	0.309		
Cond (treat):A					0.028	0.148		
Cond (treat):C					0.255	0.207		
Cond (treat):E					0.053	0.200		
TIPI(E)							-0.094	0.07
TIPI(A)							0.019	0.091
TIPI(C)							-0.159*	0.088
TIPI(ES)							0.115	0.079
TIPI(O)							0.159*	0.088
Cond (treat):TIPI(E)							0.107	0.095
Cond (treat):TIPI(A)							0.14	0.125
Cond (treat):TIPI(C)							0.256**	0.121
Cond (treat):TIPI(ES)							-0.065	0.107
Cond (treat):TIPI(O)							-0.104	0.114
Constant	-0.797***	0.136	-0.884***	0.193	-0.919***	0.226	-1.039*	0.566
Observations	1,460		1,460		1,460		1,460	
Log Likelihood	-875.25		-869.1		-846.44		-863.51	
Akaike Inf. Crit.	1,762.51		1,762.21		1,720.87		1,759.02	

Note: Regression coefficients of panel effects and interactions in (1)-(4) are not depicted for readability. Treat (1) = Loss condition. TIPI (E) = Extraversion, TIPI (A) = Agreeableness, TIPI (C) = Conscientiousness, TIPI (ES) = Emotional Stability, TIPI (O) = Openness. *p<.1; **p<.05; ***p<.01

Table S32 – Less is Better

Regression coefficients and standard errors predicting perceived generosity in the less-is-better paradigm of the regression models with panel effects (1), panel effects and demographic variables (2), panels effects and FACE factor variables (3), or panels effects and TIPI variables (4).

	Only Panels (1)		Panels & Demographics (2)		Panels & FACE (3)		Panels & TIPI (4)	
	B	SE	B	SE	B	SE	B	SE
Cond (treat)	1.107***	0.123	0.962***	0.165	1.062***	0.185	0.351	0.469
EDU			-0.110*	0.056				
Gender			-0.011	0.103				
Age			0.065	0.053				
F					0.141	0.125		
A					-0.138**	0.062		

C					0.013	0.086		
E					-0.008	0.082		
Cond (treat):F					-0.015	0.179		
Cond (treat):A					0.408***	0.085		
Cond (treat):C					0.208*	0.121		
Cond (treat):E					0.012	0.116		
TIPI(E)							-0.033	0.041
TIPI(A)							0.128**	0.055
TIPI(C)							-0.074	0.053
TIPI(ES)							0.023	0.046
TIPI(O)							0.107**	0.048
Cond (treat):TIPI(E)							-0.059	0.058
Cond (treat):TIPI(A)							0.093	0.076
Cond (treat):TIPI(C)							0.156**	0.074
Cond (treat):TIPI(ES)							-0.035	0.064
Cond (treat):TIPI(O)							-0.021	0.068
Constant	5.203***	0.087	5.122***	0.117	5.322***	0.131	4.447***	0.344
Observations	1,460		1,460		1,460		1,460	
R^2	0.204		0.216		0.248		0.236	
Adjusted R^2	0.201		0.21		0.241		0.228	

Note: Regression coefficients of panel effects and interactions in (1)-(4) are not depicted for readability. Treat (1) = Scarf condition, with the less expensive gift. TIPI (E) = Extraversion, TIPI (A) = Agreeableness, TIPI (C) = Conscientiousness, TIPI (ES) = Emotional Stability, TIPI (O) = Openness. *p<.1; **p<.05; ***p<.01

Table S33 – Sunk Cost Effect

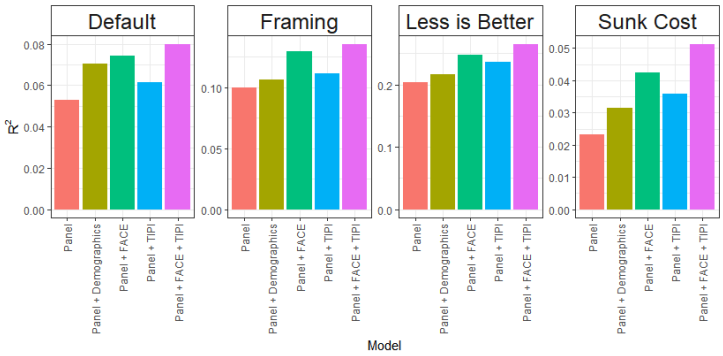
Regression coefficients and standard errors predicting self-reported likelihood to go the match in the sunk cost paradigm of the regression models with panel effects (1), panel effects and demographic variables (2), panels effects and FACE factor variables (3), or panels effects and TIPI variables (4).

	Only Panels (1)		Panels & Demographics (2)		Panels & FACE (3)		Panels & TIPI (4)	
	B	SE	B	SE	B	SE	B	SE
Cond (treat)	0.433*	0.235	0.258	0.318	0.739**	0.36	0.565	0.904
EDU			0.12	0.1				
Gender			-0.29	0.198				
Age			-0.262**	0.104				
F					0.458*	0.237		

A					0.078	0.115		
C					-0.298*	0.165		
E					-0.087	0.159		
Cond (treat):F					0.368	0.348		
Cond (treat):A					0.019	0.165		
Cond (treat):C					-0.129	0.234		
Cond (treat):E					0.15	0.226		
TIPI(E)							-0.099	0.077
TIPI(A)							0.152	0.101
TIPI(C)							-0.048	0.102
TIPI(ES)							0.098	0.089
TIPI(O)							0.193**	0.09
Cond (treat):TIPI(E)							0.138	0.112
Cond (treat):TIPI(A)							-0.121	0.147
Cond (treat):TIPI(C)							0.112	0.143
Cond (treat):TIPI(ES)							-0.097	0.125
Cond (treat):TIPI(O)						0	-0.037	0.132
Constant	5.785***	0.166	6.174***	0.225	6.002***	0.249	4.278***	0.629
Observations	1,460		1,460		1,460		1,460	
R^2	0.023		0.031		0.042		0.036	
Adjusted R^2	0.02		0.024		0.034		0.026	

Note: Regression coefficients of panel effects and interactions in (1)-(4) are not depicted for readability. Treat (1) = paid condition. TIPI (E) = Extraversion, TIPI (A) = Agreeableness, TIPI (C) = Conscientiousness, TIPI (ES) = Emotional Stability, TIPI (O) = Openness. *p<.1; **p<.05; ***p<.01

Figure S8. Predictive validity of different Moderators

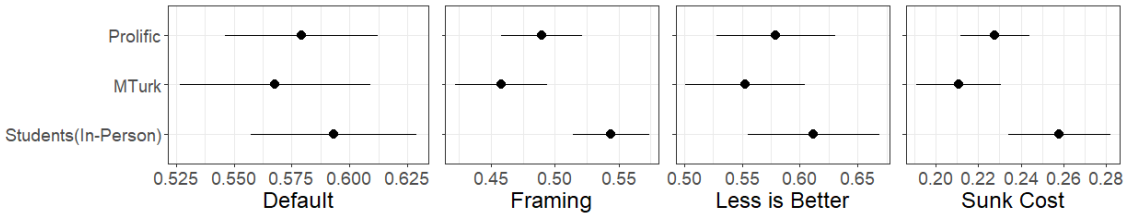


6 Survey with JDM Researchers

We recruited 67 respondents via the list server of the Society of Judgment and Decision making. Respondents received short descriptions of the Default, Framing, Less is better, and Sunk Cost paradigm together with the Cohen's d effect sizes as estimated in the Many Labs replication

study. Based on that information they were asked to estimate the effect size in an in-person student sample, and online MTurk and an online Prolific sample. Finally, we asked them to indicate their field of research, whether they used any of the three panels in the last 12 months, and how many respondents they collected in the last 12 months. 67% of respondents used Prolific, 46% used MTurk and, 40% used a student sample. Respondents reported to collect much more respondents online (Mdn = 1000, M = 4219) than students (Mdn = 0, M = 159).

Figure S9. Average Estimated Effect Sizes per Panel and Paradigm



SI References

1. H. IJzerman, *et al.*, Use caution when applying behavioural science to policy. *Nat Hum Behav* 4, 1092–1094 (2020).
2. N. Chater, G. Loewenstein, The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behav Brain Sci*, 1–60 (2022).
3. B. B. McShane, J. L. Tackett, U. Böckenholt, A. Gelman, Large-Scale Replication Projects in Contemporary Psychological Research. *Am Statistician* 73, 99–105 (2019).
4. A. Gelman, The Connection Between Varying Treatment Effects and the Crisis of Unreplicable Research. *J Manage* 41, 632–643 (2015).
5. D. A. Kenny, C. M. Judd, The Unappreciated Heterogeneity of Effect Sizes: Implications for Power, Precision, Planning of Research, and Replication. *Psychol Methods* 24, 578–589 (2019).
6. R. A. Klein, *et al.*, Investigating variation in replicability: A "many labs" replication project. *Social Psychology* 45, 142–152 (2014).
7. R. A. Klein, *et al.*, Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science* 1, 443–490 (2018).
8. T. D. Stanley, E. C. Carter, H. Doucouliagos, What Meta-Analyses Reveal About the Replicability of Psychological Research. *Psychol Bull* 144, 1325–1346 (2018).
9. K. Mrkva, N. A. Posner, C. Reeck, E. J. Johnson, Do Nudges Reduce Disparities? Choice Architecture Compensates for Low Consumer Knowledge. *J Marketing* 85, 67–84 (2021).
10. C. Ghesla, M. Grieder, R. Schubert, Nudging the poor and the rich – A field Study on the distributional effects of green electricity defaults. *Energ Econ* 86, 104616 (2020).
11. C. O. L. H. Porter, R. Outlaw, J. P. Gale, T. S. Cho, The Use of Online Panel Data in Management Research: A Review and Recommendations. *J Manage* 45, 319–344 (2019).
12. J. Chandler, D. Shapiro, Conducting Clinical Research Using Crowdsourced Convenience Samples. *Annu Rev Clin Psycho* 12, 1–29 (2015).
13. K. J. Mullinix, T. J. Leeper, J. N. Druckman, J. Freese, The Generalizability of Survey Experiments*. *J Exp Political Sci* 2, 109–138 (2015).
14. A. Krefeld-Schwalb, T. Pachur, B. Scheibehenne, Structural Parameter Interdependencies in Computational Models of Cognition. *Psychol Rev* 129, 313–339 (2022).
15. R. Baker, *et al.*, Research Synthesis: AAPOR Report on Online Panels. *Public Opin Quart* 74, 711–781 (2010).
16. M. R. Ellefson, D. M. Oppenheimer, Is Replication Possible Without Fidelity? *Psychol Methods* (2022) https:/doi.org/10.1037/met0000473.
17. C. R. Ebersole, *et al.*, Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *J Exp Soc Psychol* 67, 68–82 (2016).
18. N. Egami, E. Hartman, Elements of External Validity: Framework, Design, and Analysis. *Am Polit Sci Rev*, 1–19 (2022).

19. E. Peer, D. Rothschild, A. Gordon, Z. Evernden, E. Damer, Data quality of platforms and panels for online behavioral research. *Behav Res Methods* 54, 1643–1662 (2022).
20. A. Coppock, T. J. Leeper, K. J. Mullinix, Generalizability of heterogeneous treatment effect estimates across samples. *Proc National Acad Sci* 115, 12441–12446 (2018).
21. A. Coppock, O. A. McClellan, Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Res Politics* 6, 2053168018822174 (2019).
22. E. Snowberg, L. Yariv, Testing the Waters: Behavior across Participant Pools. *Am Econ Rev* 111, 687–719 (2021).
23. J. Chandler, G. Paolacci, E. Peer, P. Mueller, K. A. Ratliff, Using Nonnaive Participants Can Reduce Effect Sizes. *Psychol Sci* 26, 1131–1139 (2015).
24. E. Peters, Beyond Comprehension. *Curr Dir Psychol Sci* 21, 31–35 (2012).
25. G. Paolacci, J. Chandler, P. G. Ipeirotis, Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, 411–419 (2010).
26. J. K. Goodman, C. E. Cryder, A. Cheema, Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *J. Behav. Decis. Making* 26, 213–224 (2013).
27. S. Clifford, J. Jerit, Is There a Cost to Convenience? An Experimental Comparison of Data Quality in Laboratory and Online Studies. *J Exp Political Sci* 1, 120–131 (2014).
28. J. L. Horn, R. B. Cattell, Age differences in fluid and crystallized intelligence. *Acta Psychol* 26, 107–129 (1967).
29. R. B. Cattell, Theory of fluid and crystallized intelligence: A critical experiment. *J Educ Psychol* 54, 1–22 (1963).
30. N. Stewart, *et al.*, The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision Making* 10, 479–491 (2015).
31. J. K. Goodman, G. Paolacci, Crowdsourcing Consumer Research. *J Consum Res* 44, 196–210 (2017).
32. D. G. Rand, *et al.*, Social heuristics shape intuitive cooperation. *Nature Communications* 5 (2014).
33. C. R. Ebersole, *et al.*, Many Labs 5: Testing Pre-Data-Collection Peer Review as an Intervention to Increase Replicability. *Adv Methods Pract Psychological Sci* 3, 309–331 (2020).
34. E. R. Sugerman, Y. Li, E. J. Johnson, Local warming is real: A meta-analysis of the effect of recent temperature on climate change beliefs. *Curr Opin Behav Sci* 42, 121–126 (2021).
35. J. M. Jachimowicz, S. Duncan, E. U. Weber, E. J. Johnson, When and why defaults influence decisions: a meta-analysis of default effects. *Behavioural Public Policy* 3, 159–186 (2019).
36. Y. Li, E. J. Johnson, L. Zaval, Local Warming. *Psychol Sci* 22, 454–459 (2010).
37. E. J. Johnson, D. Goldstein, Do Defaults Save Lives? *Science* 302, 1338–1339 (2003).
38. A. Tversky, D. Kahneman, “Judgments of and by Representativeness” (1981).
39. C. K. Hsee, Less is better: when low-value options are valued more highly than high-value options. *J. Behav. Decis. Making* 11, 107–121 (1998).
40. D. M. Oppenheimer, T. Meyvis, N. Davidenko, Instructional manipulation checks: Detecting satisficing to increase statistical power. *J Exp Soc Psychol* 45, 867–872 (2009).
41. L. Ross, D. Greene, P. House, The “false consensus effect”: An egocentric bias in social perception and attribution processes. *J Exp Soc Psychol* 13, 279–301 (1977).
42. M. Hauser, F. Cushman, L. Young, R. K.-X. Jin, J. Mikhail, A Dissociation Between Moral Judgments and Justifications. *Mind Lang* 22, 1–21 (2007).
43. T. Dohmen, *et al.*, Individual Risk Attitudes: Measurement, Determinants, And Behavioral Consequences. *J Eur Econ Assoc* 9, 522–550 (2011).
44. R. Frey, A. Pedroni, R. Mata, J. Rieskamp, R. Hertwig, Risk preference shares the psychometric structure of major psychological traits. *Sci Adv* 3, e1701381 (2017).
45. I. W. Eisenberg, *et al.*, Uncovering the structure of self-regulation through data-driven ontology discovery. *Nat Commun* 10, 2319 (2019).
46. J. Chapman, M. Dean, P. Ortoleva, E. Snowberg, C. Camerer, Econographics. *J Political Econ Microeconomics* 1, 115–161 (2023).
47. A. Gelman, J. Lax, J. Phillips, J. Gabry, R. Trangucci, Using Multilevel Regression and Poststratification to Estimate Dynamic Public Opinion (2018).