SUPPLEMENTAL MATERIALS
TABLE OF CONTENTS

I. SUPPLEMENTAL METHOD FOR PRIMARY STUDIES

Studies 1-3: Validating the Empathy Selection Task

Sample Information. A sensitivity analysis suggests that with 50 participants and our within-subject design, we have 80% power to detect effect sizes of d = .40 or larger. Given that the average effect size in social psychology is d = .40, we would achieve 97% power with 100 people and 79% power with 50 people. We thus aimed to recruit at least 50 participants per study, but sometimes closer to 100 per study. In Study 1, we enrolled 133 MTurk participants. From this initial sample, 77 participants dropped out before finishing the Empathy Selection Task, leaving a final sample of 56 participants (29 female, 27 male, M_{age} = 38.36, SD_{age} = 12.22). In Study 2, we enrolled 134 MTurk participants. From this initial sample, 83 participants dropped out before finishing the Empathy Selection Task. We also excluded 3 participants who skipped at least one response on the Empathy Selection Task, and 1 participant who had completed an earlier study in this sequence, leaving a final sample of 47 participants (29 female, 18 male, M_{age} = 40.45, SD_{age} = 12.66). In Study 3, we enrolled 260 MTurk participants. From this initial sample, 60 participants dropped out before finishing the Empathy Selection Task. We also excluded 3 participants who skipped at least one response on the Empathy Selection Task, and 1 participant who had completed an earlier study in this sequence, leaving a final sample of 196 participants (111 female, 85 male, M_{age} = 36.60, SD_{age} = 10.76). Subject attrition on Amazon.com Mechanical Turk occurs frequently, and we excluded participants who skipped responses on the logic that they may not have been fully engaged with the task.

Empathy Selection Task.

Study 1. In the pre-task instructions, participants were told:

> In this task, you will complete a series of trials. On each trial, you will see two decks of cards: the deck on the left will always be labeled "DESCRIBE" and the deck on the right will always be labeled "FEEL." You should choose between these decks. Once you choose a deck, you will then see an image of a person. Depending on which deck you have chosen, you will be given one of two possible sets of instructions.
>
> If you choose from the deck labeled "DESCRIBE", you will be told to be objective and focus on the external features and appearances of the person in the image. When completing this kind of trial, try to be as objective as possible. To be objective, do not let yourself get caught up in imagining what this person feels. On these trials, describe the age and gender of the person.
>
> If you choose from the deck labeled "FEEL", you will be told to have empathy and focus on the internal feelings and experiences of the person in the image. When completing this kind of trial, try to feel as much empathy as possible. To be empathic, let yourself get caught up in imagining what this person feels. On these trials, describe the feelings and experiences of the person.

> You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. Overall, this task will take the same amount of time regardless of which deck you choose.

At the beginning of each trial, participants were shown a pair of card decks. The deck on the left was red and labeled "DESCRIBE" and the deck on the right was blue and labeled "FEEL." There was not a time limit on choice. Once a choice was made participants saw an image of a child refugee. Refugee images were selected from online media, and were randomized and not repeated. If participants chose DESCRIBE, they were instructed: "Look at the person in the picture, and **try to notice details about this person**. Objectively focus on the external features and appearance of this person. Please write one sentence describing the age and gender of this person." If participants chose FEEL, they were instructed: "Look at the person in the picture, and **try to feel what this person is feeling**. Empathically focus on the internal experiences and feelings of this person. Please write one sentence describing the experiences and feelings of this person." A Qualtrics timer was incorporated on the written responses so that participants could not submit a response until after 10 seconds had elapsed. Participants completed 40 trials which were presented in randomized order. In Studies 1-2 and 4-5, on four of the trials the trial-level instruction said to "look at the hand in the picture" instead of "look at the person in the picture"; however, we do not believe that such rare typos influenced task performance, given that choice preferences aggregated across a large number of trials.

Study 2. Task instructions and details were identical to Study 1, except that the decks were labeled "DECK 1" and "DECK 2" rather than "FEEL" and "DESCRIBE" in the Empathy Selection Task as well as in the post-task assessment. The pre-task instruction removed reference to deck labels, instead saying "On some trials…" and "On other trials…"

Study 3. The task instructions were different in Study 3, requiring different kinds of information to be provided. Participants received the following pre-task instructions:

> In this task, you will complete a series of trials. On each trial, you will see two decks of cards. You should choose between these decks. Once you choose a deck, you will then see an image of a person. These are actors that we've asked to look certain ways and express certain emotions. Depending on which deck you have chosen, you will then be given one of two possible sets of instructions.
>
> On some trials, you will be told to be objective and focus on identifying the emotional expression of the person in the image. When completing this kind of trial, try to be objective. To be objective, try to focus on which emotions the person's facial expression most closely resembles. On these trials, please provide three keywords to describe the facial expression of the person (Example: "sad, hurt, confused" or "happy, pleased, interested"). It is okay to use the same keywords multiple times, just make sure that you are describing the emotional expression of the person in the image.
>
> On other trials, you will be told to have empathy and share in the emotional experience of the person in the image. When completing this kind of trial, try to feel empathy. To be

empathic, try to share in the internal emotional experience of the person. On these trials, please provide three keywords to describe the emotional experience of this person (Example: "sad, hurt, confused" or "happy, pleased, interested"). It is okay to use the same keyword multiple times, just make sure you are describing the feelings and experiences of the person in the image.

You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. Overall, this task will take the same amount of time regardless of which deck you choose.

To ensure that participants were comprehending the revised instructions, participants were asked: "Which of the following is an appropriate response on trials where you are told to be objective? (Be emotional, and provide keywords describing the internal emotional experiences you are feeling; Be objective, and provide keywords describing the facial emotional expression of the person; Be empathic, and provide keywords describing the internal emotional experience of the person)" Participants were also answered the question "Which of the following is an appropriate response on trials where you are told be empathic?" with the same response options. Participants could not proceed until they had provided the correct answers.

On objective trials, participants were instructed: "Look at the person in the picture, and **try to identify the emotion of this person**. Objectively focus on the external facial expression of this person. Please write 3 keywords describing the objective facial expression of this person." On empathy trials, participants were instructed: "Look at the person in the picture, and **try to feel what this person feels**. Empathically share in the internal emotional experience of this person. Please write 3 keywords describing the subjective emotional experience of this person." Participants completed 40 trials in randomized order. The target images were 40 unique Black/White female/male actors from the Chicago Face Database, all displaying anger (Ma, Correll, & Wittenbrink, 2015; Black female exemplars: BF001, BF008, BF021, BF031, BF037, BF040, BF047, BF048, BF042, BF050; Black male: BM002, BM011, BM013, BM018, BM020, BM024, BM031, BM032, BM037, BM046; White female: WF001, WF006, WF009, WF020, WF030, WF031, WF033, WF035, WF037, WF039; White male: WM006, WM009, WM013, WM014, WM015, WM025, WM028, WM032, WM033, WM040).

Post-Task Questionnaire. Participants provided open-ended responses to the following: "What was it like performing the task?" "How did you choose between decks?" "Did you develop a preference for one of the decks?" "Was there any difference between the decks?"

NASA Task Load Index. Participants completed a version of the NASA Task Load Index (Hart & Staveland, 1988). They rated each Empathy Selection Task deck on the following questions (from 1=*Very low to* 5=*Very high*): "How mentally demanding was this deck?" "How hard did you have to work to accomplish your level of performance with this deck?" "How insecure, discouraged, irritated, stressed, and annoyed were you by this deck?" "How successful were you in accomplishing what you were asked to do in this deck?" Participants rated the objective deck first, which was referred to as "the Describe deck (the one on the left)", followed

by the objective deck, referred to as "the Feel deck (the one on the right)". Decks were referred to as "Deck 1" and "Deck 2" in Study 2, and "Deck A" and "Deck B" in Study 3.

Additional Measures. In Study 3, participants completed two measures of trait empathy: the 28-item Interpersonal Reactivity Index (Davis, 1983), and the 14-item Empathy Index (Jordan, Amir, & Bloom, 2016). In Study 3, participants also completed measures of social norms about empathy. Participants were told: "Please answer the following questions, as honestly as possible, based on your own personal opinion and experiences." First, they were asked about descriptive norms of choice in the Empathy Selection Task (scale from 0-100): "What percentage of people tend to choose the empathy deck in the task you just completed?" "What percentage of people tend to choose the objective deck in the task you just completed?" Second, they were asked about injunctive norms about empathy and objectivity (scale from 0-100): "What percentage of people think that empathy is a good thing?" "What percentage of people think that objectivity is a good thing?" Finally, participants were asked: "According to your own personal beliefs, do you think that it is more desirable to be empathic or to be objective" (1=*objectivity is much more desirable,* 2=*objectivity is more desirable*, 3=*they are equally desirable*, 4=*empathy is more desirable*, 5=*empathy is much more desirable*). Finally, in all studies, participants reported gender, age, political orientation (from 1=*Extremely liberal* to 7=*Extremely conservative*), religiosity (from 1=*Not at all religious* to 5=*Extremely religious*), and socioeconomic status using the MacArthur ladder (Adler & Ostrom, 1999).

Studies 4-6: Manipulating Valence of Target Affect

Sample Information. In Study 4, we enrolled 359 MTurk participants. From this initial sample, 158 participants dropped out before finishing the Empathy Selection Task. Of those who dropped out, 38 did not progress far enough to receive randomizer information, 61 were in the negative condition, and 58 were in the positive condition. We also excluded 4 participants who skipped at least one response on the EST, and 4 participants who had completed an earlier study in this sequence, leaving a final sample of 193 participants (108 female, 85 male, M_{age} = 36.73, SD_{age} = 11.63). In Study 5, we enrolled 326 MTurk participants. From this initial sample, 115 participants dropped out before finishing the Empathy Selection Task. Of those who dropped out, 10 did not progress far enough to receive randomizer information, 47 were in the negative condition, and 58 were in the positive condition. We also excluded 4 participants who skipped at least one response on the EST, and 1 participant who had completed an earlier study in this sequence, leaving a final sample of 206 participants (117 female, 83 male, 6 unreported, M_{age} = 36.48, SD_{age} = 12.16). In Study 6, we enrolled 80 MTurk participants. From this initial sample, 29 participants dropped out before finishing the EST. We also excluded 1 participant who skipped at least one response on the EST, leaving a final sample of 50 participants (28 female, 22 male, M_{age} = 35.14, SD_{age} = 9.67).

Empathy Selection Task.

Studies 4-5. In these studies, the task was nearly identical to Study 1, except that there was a between-subjects manipulation of target valence (negative, positive), and the target images were changed from child refugees to college-age adults. Target images were the same Chicago Face Database actors from Study 3. In the negative condition, these actors displayed anger; in

the positive condition, these actors displayed happiness. Target images were presented in random order. In Study 4, there was a programming error on one of the trials in the positive condition: participants did not see a target image to write a sentence about. This typo was fixed in Study 5. Because preferences emerge across 40 trials, we do not believe that this rare typo significantly influenced choice behavior. In Study 5, the task was identical to Study 4, except decks were unlabeled (i.e., "DECK 1" and "DECK 2").

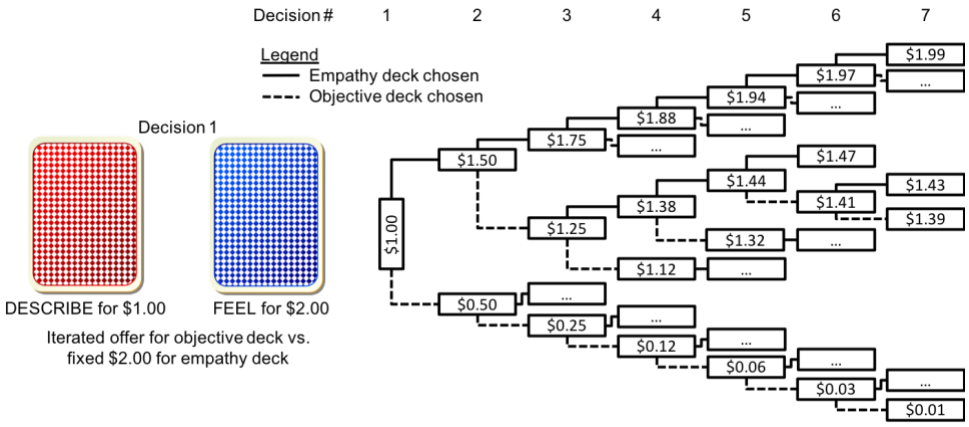
Study 6. The Empathy Selection Task was nearly identical to Study 3, except that targets were presented prior to choices and valence was manipulated within-subjects. Pre-task instructions were nearly identical to Study 3. Participants received a comprehension check as in Study 3. Trial-level instructions were identical to Study 3. Participants completed 40 trials in randomized order. Target images were 20 White female and male actors from the NimStim Database, presented once displaying happiness and once displaying sadness (Tottenham et al., 2009; 01F, 02F, 03F, 05F, 06F, 07F, 08F, 09F, 10F, 20M, 21M, 22M, 23M, 24M, 25M, 26M, 28M, 29M, 30M, 31M). On each trial, the target was presented on screen for 5 seconds, before participants then saw two bullets labeled "DECK A" and "DECK B" underneath the target, with no card deck images displayed.

Post-Task Questionnaire. This was identical to previous studies.

NASA Task Load Index. Participants completed the same questions as in Studies 1-3.

Empathy Discounting Paradigm. In Study 5, participants completed an additional task (modeled on Westbrook et al., 2013), after the Empathy Selection Task, in order to measure subjective value assigned to empathy:

> In the next part of the experiment, you will be making choices between hypothetical amounts of money. You will see two decks of cards, exactly like the task you completed earlier. The Describe deck asks you to be objective and write about the age and race of a person, and the Feel deck asks you to be empathic and write about the internal experiences and feelings of a person. In all cases, the persons shown will be crying children similar to those you saw earlier in the experiment. This time, you will see a hypothetical payment for completing a trial from each deck. Please select the option that you prefer. There are no accurate or inaccurate answers. Unlike the earlier part of the experiment, you will not have to write about a person after you make the choice.



The task schematic (Figure 7 in main text) is displayed above (modeled on Westbrook et al., 2013). Dollar values indicate iterated cost of objective deck on each trial, depending on choice on previous trial. Participants make a series of choices between completing a trial from the objective deck for a varying lesser amount, or a trial from the empathy deck for a fixed larger amount ($2.00). Participants do not complete the trial after the choice, and payment is stipulated as hypothetical. If the larger (smaller) offer is selected, the offer for the objective deck is increased (decreased) on subsequent choice. The amount of increase/decrease halves with each choice, such that adjustments following the seventh choice are $0.0075, with the final value after that adjustment (rounded to the nearest cent) reflecting the point of indifference between the decks. Subjective cost of empathy is computed as offer for the empathy deck ($2.00) minus the indifference point, quantifying additional money to empathize.

Additional Measures. Demographics were assessed as in previous studies.

Studies 7-8 (Feel-Self vs. Feel-Other Variant)

Sample Information. In Study 7, we enrolled 118 MTurk participants. From this initial sample, 19 participants dropped out before finishing the Empathy Selection Task. We also excluded 2 participants who skipped at least one response on the Empathy Selection Task, and 6 participants who had completed an earlier study in this sequence, leaving a final sample of 91 participants (39 female, 51 male, 1 other, M_{age} = 35.31 years, SD_{age} = 12.41 years). In Study 8, we enrolled 98 MTurk participants. We excluded 4 participants who skipped at least one response on the Empathy Selection Task, and 7 participants who had completed an earlier study in this sequence, leaving a final sample of 87 participants (46 female, 41 male, M_{age} = 38.28 years, SD_{age} = 12.54 years).

Empathy Selection Task. As noted in the primary text, this variant of the Empathy Selection Task was structured differently. First, participants received the following pre-task instructions (the sentence about explicit content was not present in Study 8):

> In this task, you will complete a series of trials. On each trial, you will see two decks of cards. You should choose between these decks. Once you choose a deck, you will then see an image. **Please note that some of these images contain sexually explicit content.** Depending on which deck you have chosen, you will be given one of two possible sets of instructions.

> On some trials, you will be asked to reflect on the emotional reactions **that you are having to the image.** When completing this kind of trial, try to focus on the emotions and feelings you are experiencing in response to the image. Please sit with each image for at least 10 seconds, and let yourself get caught up in how you feel. Then indicate whether you feel more positive or negative overall. **Note: Sometimes you may feel conflicted, or may not feel much at all. That's okay. Simply attend to how you are feeling and the indicate whether you feel more positive or negative overall.**
>
> On other trials, you will be asked to reflect on the emotional reactions **that another participant is having to the image**. When completing this kind of trial, try to focus on the emotions and feelings that you expect the other participant is experiencing in response to the image. Please sit with each image for at least 10 seconds, and let yourself get caught up imagining what the other participant feels. Then indicate whether they feel more positive or negative overall. **You may feel that they are conflicted, or don't experience much at all, but you should still just indicate whether the image makes them feel more positive or negative overall.**
>
> You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. **Overall, this task will take the same amount of time regardless of which deck you choose.**

To enhance the believability that another person was completing the task, before starting the task participants entered their first name, and the day and month of their birth (e.g., "Sam, June 21"). At the beginning of each trial of the task, participants were shown a pair of card decks. The deck on the left was labeled "FEEL-SELF" and the deck on the right was labeled "FEEL-OTHER." Once a choice was made participants saw a slide from the International Affect Picture System (IAPS). In Study 7, these consisted of positive and negative low-arousal images presented in randomized order (image IDs: 1303, 1935, 2383, 2393, 2440, 2441, 2485, 2493, 2514, 2518, 2579, 2595, 2690, 2702, 2749, 2780, 3550.2, 4000, 4003, 4005, 4275, 4571, 5390, 6930, 7002, 7004, 7030, 7035, 7096, 7150, 7160, 7182, 7185, 7235, 7285, 7504, 7550, 7590, 7820, 7830). In Study 8, the IAPS slides were replaced with medium-arousal images (image IDs: 1051, 1220, 1650, 1660, 2110, 2515, 2560, 2630, 2655, 2682, 2795, 3022, 4150, 4220, 4533, 4598, 4689, 5622, 5990, 5994, 6010, 6200, 6211, 6250.2, 6610, 6940, 7220, 7281, 7360, 8041, 8280, 8480, 9090, 9101, 9156, 9190, 9230, 9341, 9373, 9404). An avatar was displayed beneath the IAPS slide. If participants chose Feel-Other, they saw "Reilly, July 22" and were instructed: "Look at the picture, and **focus on what emotional reactions Reilly is having to the picture**. How is Reilly feeling right now?" If participants chose FEEL-SELF, they saw their own information and were instructed: "Look at the picture, and **focus on what emotional reactions you are having to the picture**. How are you feeling right now?" Participants made a binary response by selecting "POSITIVE" or "NEGATIVE." Participants completed 40 trials in randomized order. To account for stimulus sampling, in Study 8 half of participants saw "Harley, October 3" and half saw "Casey, January 14." The names Reilly, Harley, and Casey were chosen as gender-neutral.

Post-Task Questionnaire. This was identical to previous studies.

NASA Task Load Index. Participants completed the same questions as in earlier studies.

Additional Measures. Participants completed the IRI, the 20-item Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994) and the 20-item Philadelphia Mindfulness Scale (Cardaciotto et al., 2008). Demographics were assessed as in previous studies.

Studies 9-10 (Efficacy Manipulation)

Sample Information. In Study 9, we enrolled 134 MTurk participants. From this initial sample, 34 participants dropped out before finishing the Empathy Selection Task. Of these participants who dropped out, 3 did not progress far enough to receive randomizer information, 18 were in the low-efficacy condition, and 13 were in the high-efficacy condition. We excluded 1 participant who skipped at least one response on the Empathy Selection Task, and 9 participants who had completed an earlier study in this sequence, leaving a final sample of 90 participants (50 female, 40 male, M_{age} = 34.51 years, SD_{age} = 10.13 years). In Study 10, we enrolled 116 MTurk participants. From this initial sample, 16 participants dropped out before finishing the Empathy Selection Task. Of these participants who dropped out, 10 were in the low-efficacy condition and 6 were in the high-efficacy condition. We excluded 7 participants who had completed an earlier study in this sequence, leaving a final sample of 93 participants (57 female, 36 male, M_{age} = 37.64 years, SD_{age} = 12.11 years).

Efficacy Manipulation & Empathy Selection Task. We measured perceived efficacy of empathy and emotion self-awareness before and after training trials of the Empathy Selection Task (on a sliding scale from 0=*Not good at all* to 100=*Incredibly good*): "I usually feel like I am very aware of and good at understanding exactly what I'm feeling." "I usually feel like I am very aware of and good at understanding exactly what other people are feeling." In Study 9, but not Study 10, participants read the following paragraph prior to the training phase:

> You are about to complete a task that measures **empathy**. People who are high in empathy are especially good at figuring out how the people around them are feeling. This is a good skill to have—for example, research has shown that people high in empathy are viewed more positively by their peers, are more likely to succeed in the workplace, and tend to have stronger personal relationships. All in all, empathy is viewed as a good/useful skill, and one that most people want to have.

Participants proceeded to instructions for the Empathy Selection Task:

> In this task, you will complete a series of trials. On each trial, you will see two decks of cards. You should choose between these decks. Once you choose a deck, you will then see an image of a person. These are actors that we've asked to look certain ways and express certain emotions. Depending on which deck you have chosen, you will be given one of two possible sets of instructions.

> On the DESCRIBE trials, you will be told to be objective and focus on the external features and appearances of the person in the image. When completing this kind of trial, try to be as objective as possible. To be objective, do not let yourself get caught up in imagining what this person feels. On these trials, please provide three keywords to describe the physical appearance of the person, as if you were describing them to a sketch artist. (Example: "white, woman, long hair" or "young, black, blue eyes"). It is ok to use the same keywords multiple times, just make sure that you are accurately describing the physical appearance of the person in the image (e.g., age, gender, race, etc.)
>
> On the FEEL trials, you will be told to have empathy and focus on the internal feelings and experiences of the person in the image. When completing this kind of trial, try to feel as much empathy as possible. To be empathic, let yourself get caught up in imagining what this person feels. On these trials, please provide three keywords to describe the feelings and experiences of the person. (Example: "sad, hurt, confused" or "happy, pleased, interested"). It is ok to use the same keywords multiple times, just make sure that you are a describing the internal feelings and experiences of the person in the image.
>
> You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. Overall, this task will take the same amount of time regardless of which deck you choose.

To ensure that participants were comprehending the instructions, they were asked: "Which of the following is an appropriate response for the DESCRIBE deck trials? (smiling, woman, white; woman, young, blond; happy, smiling, blond)" "Which of the following is an appropriate response for the FEEL deck trials? (smiling, white, woman; woman, happy, blond; happy, interested, curious)". Participants could not proceed until they had provided the correct answer.

The efficacy manipulation was embedded into a subsequent training phase. Participants were informed: "You will complete a set of practice trials. These trials are meant to give you experience with both the DESCRIBE and FEEL decks, so you will complete each separately." Participants completed practice blocks separately for the DESCRIBE and FEEL decks, in counterbalanced order. For the DESCRIBE practice, participants read the following:

> This is your DESCRIBE practice. On the DESCRIBE trials, you will be told to be objective and **focus on the external features and appearances of the person in the image**. When completing this kind of trial, try to be as objective as possible. To be objective, do not let yourself get caught up in imagining what this person feels. On these trials, please provide three keywords to **describe the physical appearance of the person**, as if you were describing them to a sketch artist. (Example: "white, woman, long hair" or "young, black, blue eyes"). It is ok to use the same keywords multiple times, just make sure that you are accurately describing the physical appearance of the person in the image (e.g., age, gender, race, etc.) **For each image, a panel of judges has identified the most descriptive and helpful keywords for distinguishing the person in a sketch. If at least 2 of your keywords match the judges' keywords, we'll let you know during the practice trials!**

Participants then completed four modified trials of the Empathy Selection Task. In the DESCRIBE practice, only the DESCRIBE deck was presented for participants to select. After making this selection, participants saw each of four White female exemplars from the Chicago Face Database (WF033, WF035, WF037, WF039), with two expressing anger and two expressing happiness, and were instructed:

> Look at the person in the picture, and **try to notice details about this person**. Objectively focus on the external features and appearance of this person. Please provide 3 keywords describing the objective physical features of this person.

For the FEEL practice, participants read the following:

> This is your FEEL practice. Remember: On the FEEL trials, you will be told to **have empathy and focus on the internal feelings and experiences of the person in the image**. When completing this kind of trial, try to feel as much empathy as possible. To be empathic, let yourself get caught up in imagining what this person feels. On these trials, please provide three keywords to **describe the feelings and experiences of the person**. (Example: "sad, hurt, confused" or "happy, pleased, interested"). It is ok to use the same keywords multiple times, just make sure that you are a describing the internal feelings and experiences of the person in the image. **We had the actors write down several keywords describing their own experience. If at least 2 of your keywords match their list of keywords, we'll let you know during the practice trials!**

As in the DESCRIBE practice, participants completed four FEEL trials. Targets for the FEEL trials were the same as in the DESCRIBE trials, but displaying the other emotion (i.e., anger instead of happiness, or vice versa). On each trial, participants were instructed:

> Look at the person in the picture, and **try to feel what this person is feeling**. Empathically focus on the internal experiences and feelings of this person. Please write 3 keywords describing the experiences and feelings of this person.

After making a response, participants saw their response displayed back to them along with accuracy feedback. On accurate-feedback trials, participants read: "Correct! Your keywords matched." On inaccurate-feedback trials, participants read: "Incorrect. No keyword match." In the low-efficacy condition, participants received feedback as accurate on all DESCRIBE practice trials and on half of the FEEL practice trials. After DESCRIBE practice, these participants read:

> Of all the people tested so far, you scored in the: top 5th percentile of all Workers for the DESCRIBE deck practice trials! This means that you were better at describing the physical appearance of the person in the image than 95 percent of people who have completed the DESCRIBE deck practice trials. Congratulations!

After the FEEL practice, the low-efficacy participants read:

> Of all the people tested so far, you scored in the: top 50th percentile of all Workers for the FEEL deck practice trials! This means that you were better at knowing how the actor felt than 50 percent of people who have completed the FEEL deck practice trials.

In the high-efficacy condition, feedback was reversed. Participants received feedback that they were accurate on 100% of the FEEL practice trials and 50% of the DESCRIBE practice trials. Participants received the 5th percentile feedback after the FEEL practice and the 50th percentile feedback after the DESCRIBE practice. After this efficacy manipulation, participants completed the same questions about efficacy of empathy and of emotion self-awareness. Participants then completed the NASA Task Load Index, as in prior studies. Before proceeding to test trials of the Empathy Selection Task, participants were shown the pre-task instructions again as a reminder. Trial-level instructions were identical to the training phase, except that now participants could choose between the FEEL and DESCRIBE decks, and no feedback was provided after each trial. In the test trials, the exemplars were actors from the Chicago Face Database, with each actor presented once displaying anger and once displaying a happiness (BF001, BF008, BF021, BF031, BF037, BF050, WF001, WF006, WF009, WF020, WF030, WF031). Targets were randomized across trials. Participants completed 24 trials of the task.

Post-Task Questionnaire. This was identical to previous studies.

NASA Task Load Index. Participants completed identical questions from earlier.

Funneled Debriefing. Participants completed a funneled debriefing by providing open-ended responses to the following questions: "What did you think of this study?" "What did you think the purpose of the study was?" "Were you suspicious about anything?" "Did you think we were trying to deceive you?" "During the practice trials, you were provided with feedback about your performance on the DESCRIBE deck (the red one). What did you think about that feedback?" "During the practice trials, you were provided with feedback about your performance on the FEEL deck (the blue one). What did you think about that feedback?" Lastly, participants were asked "Did you think the feedback that we gave you was accurate?" (Yes, I did; No, I didn't; I didn't think about it much; I wasn't sure, but I was skeptical).

Additional Measures. We assessed trait empathy on the IRI and demographics. In Study 10, participants were asked "What were you thinking about when you were deciding which deck to choose?" Three questions were presented (from 1=*Strongly disagree* to 7=*Strongly agree*): "I wanted to show how much empathy I have", "I wanted to practice my empathy skills", and "I wanted to avoid feeling empathy." Participants also completed an attention check: "Did you pay attention while completing this study? Did you read the instructions and complete the task as requested? Your answer here WILL NOT affect your payment?" (Yes, I paid full attention. You should use my data; No, I didn't pay close attention. You should not use my data.) All participants indicated paying attention.

Study 11 (Varying Empathic Demand)

Sample Information. In Study 11, we enrolled 315 MTurk participants. From this initial sample, 112 participants dropped out before finishing the Empathy Selection Task. We also

excluded 2 participants who skipped at least one written response on the Empathy Selection Task, 3 participants who entered nonsensical responses that indicated not following task instructions, and 5 participants who had completed an earlier study in this sequence, leaving a final sample of 193 participants (109 female, 83 male, 1 unreported, M_{age} = 38.78 years, SD_{age} = 12.69 years). From this final sample, 1 participant skipped state upset ratings on the Empathy Selection Task, and 42 participants indicated that they did not follow the instruction to empathize for the full amount of time on at least one post-trial check question (see details below). To maximize statistical power, we opted to retain these participants in the sample for analyses, but excluding them does not change results.

Empathy Selection Task. Participants completed a modified variant of the Empathy Selection Task, receiving the following pre-task instructions:

> In this task, you will complete a series of trials. On each trial, you will see two decks of cards: one deck will be a red deck labeled "FEEL-3" and the other will be a blue deck labeled "FEEL-10". You should choose between these decks. Once you choose a deck, you will then see an image of a person. **Depending on which deck you choose, you will be asked to empathize for different amounts of time: either 3 seconds or 10 seconds.**
>
> **When you see the instruction "EMPATHIZE" on the screen**, try to feel as much empathy as possible for this person. To be empathic, try to share in the person's suffering, feelings, and experiences. **When you don't see the instruction "EMPATHIZE" on the screen**, you can stop trying to empathize. In other words, only try to feel empathy when you see the instruction to do so on screen.
>
> At the end of each trial, you will be asked to enter 3 keywords to describe the person's feelings and internal emotional experiences (Example: "sad, hurt, confused" or "happy, pleased, interested"). It is okay to use the same keyword multiple times, just make sure you are describing the feelings and experiences of the person in the image (e.g., mood, emotion, etc.).
>
> You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. Overall, this task will take the same amount of time regardless of which deck you choose. Note also that the decks will switch sides during the course of the task.

At the beginning of each trial of the task, participants were shown a pair of card decks. There was a red deck labeled “FEEL-3” and a blue deck labeled “FEEL-10”. The positioning of each deck on left vs. right was randomly counterbalanced across trials. As in previous studies, there was not a time limit on choice. Once a choice was made participants saw an image of a child refugee, as in Studies 1-2 and 4-5. For both deck choices, once participants made a choice they were instructed: “Look at the person in the picture. EMPATHIZE. **Try to feel what this person feels.**” If participants selected the FEEL-3 deck, they viewed the refugee and empathy instruction for three seconds; if they selected the FEEL-10 deck, they viewed these for ten seconds. For both deck choices, after the empathy instruction screen they viewed the image again and were

instructed "Please write 3 keywords describing the feelings and experiences of this person." A Qualtrics timer was incorporated on the written responses so that participants could not submit a response until after 5 seconds had elapsed. After entering a written response, participants completed a state empathy rating: "How upset are you for the child you just saw?" (from 1=*not at all upset* to 9=*extremely upset*), paired with a self-assessment manikin image (Bradley & Lang, 1994). Finally, the last component of the trial was a question asking participants to indicate time spent on task. Participants were asked "Did you work at your empathy task for the full 3 [10] seconds? If not, how many seconds did try you empathize?" The number of the seconds mentioned in the question was contingent on their choice, as a means of further reinforcing the difference between the decks. Due to a programming glitch, on one of the trials, participants completed both versions of this question. Participants could answer Yes or No, and were instructed to specify number of seconds if they answered no. Participants completed 40 trials which were presented in randomized order.

Post-Task Questionnaire. This was identical to previous studies.

NASA Task Load Index. Participants completed identical questions from earlier. In addition to the questions about effort, aversiveness, and efficacy, participants also completed three new questions per deck: "How emotionally rewarding was this deck?" "How socially rewarding was this deck?" "How valuable did you find this deck?"

Donation. Participants were asked: "How much would you be willing to donate to Save the Children, an international relief organization?" Scale anchors ranged from $0.00 to $5.00 in increments of $0.50, with a Qualtrics slider that allowed participants to move in increments of $0.01.

Additional Measures. We assessed trait empathy on the IRI and Empathy Index, the Fear of Compassion for Others Scale, and demographics.

II. SUPPLEMENTAL RESULTS FOR PRIMARY STUDIES

Table S1. NASA Task Load Index effort ratings, Studies 1-11.

Study	Emp-Obj M_{diff} [95% CI]	*p*	Emp-Obj Hedges' *g*	Choice *r*	*p* for Choice *r*	*N*
1. Labeled decks	0.99 [0.66, 1.32]	< .001	0.94	-.20	.148	56
2. Non-labeled decks	1.01 [0.73, 1.30]	< .001	0.90	-.32	.029	47
3. Empathy vs. emotion recog.	0.20 [0.09, 0.32]	< .001	0.17	-.26	< .001	196
4. Valence manip., between	0.60 [0.43, 0.77]	< .001	0.54	-.22	.002	193
5. Valence manip., between	0.70 [0.54, 0.86]	< .001	0.61	-.28	< .001	206
6. Valence manip., within	-0.02 [-0.31, 0.27]	.891	-0.02	.09	.525	50
7. Feel self/other, low arousal	0.46 [0.25, 0.66]	< .001	0.38	-.11	.291	91
8. Feel self/other, med. arousal	0.48 [0.26, 0.70]	< .001	0.42	-.21	.057	87
9. Efficacy manipulation	1.22 [0.85, 1.59]	< .001	1.11	-.42	< .001	90
10. Efficacy manipulation	1.77 [1.42, 2.12]	< .001	1.89	-.45	< .001	93
11. Varying empathic demand	0.14 [0.04, 0.24]	.007	0.13	-.04	.604	192
Meta-analytic effect		< .001	**0.56**	**-.23**	< .001	1301

Note on Tables S1-S3. Mean difference computed as rating for empathy deck minus rating for objective deck. For deck comparisons, the low-efficacy conditions in Studies 9-10 are excluded because typical deck differences in these costs were expected to be reduced (N = 1203 for effort and aversion, N = 1202 for efficacy). For associations with empathy choice, these conditions are re-included because associations between choice and effort costs were expected to be similar across efficacy conditions (N = 1301 for effort and aversion, N = 1300 for efficacy). Values in the sample size column correspond to the analysis including high-efficacy conditions. The bottom row values are from random-effects meta-analyses.

Table S2. NASA Task Load Index aversion ratings, Studies 1-11.

Study	Emp-Obj M_{diff} [95% CI]	*p*	Emp-Obj Hedges' *g*	Choice *r*	*p* for Choice *r*	*N*
1. Labeled decks	0.71 [0.33, 1.10]	< .001	0.54	-.11	.403	56
2. Non-labeled decks	0.94 [0.56, 1.32]	< .001	0.72	-.16	.283	47
3. Empathy vs. emotion recog.	0.16 [0.03, 0.29]	.018	0.12	-.35	< .001	196
4. Valence manip., between	0.47 [0.30, 0.64]	< .001	0.33	-.17	.018	193
5. Valence manip., between	0.39 [0.22, 0.56]	< .001	0.28	-.24	< .001	206
6. Valence manip., within	0.12 [-0.22, 0.46]	.485	0.09	-.14	.316	50
7. Feel self/other, low arousal	0.44 [0.23, 0.65]	< .001	0.33	-.12	.244	91
8. Feel self/other, med. arousal	0.22 [-0.01, 0.44]	.058	0.19	-.18	.105	87
9. Efficacy manipulation	1.22 [0.78, 1.66]	< .001	1.19	-.55	< .001	90
10. Efficacy manipulation	1.33 [1.01, 1.65]	< .001	1.33	-.35	.001	93
11. Varying empathic demand	0.04 [-0.13, 0.21]	.675	0.03	-.08	.282	192
Meta-analytic effect		< .001	**0.37**	**-.23**	< .001	1301

Table S3. NASA Task Load Index efficacy ratings, Studies 1-11.

Study	Emp-Obj M_{diff} [95% CI]	*p*	Emp-Obj Hedges' *g*	Choice *r*	*p* for Choice *r*	*N*
1. Labeled decks	-0.38 [-0.74, -0.01]	.043	-0.34	.26	.056	56
2. Non-labeled decks	-0.74 [-1.08, -0.41]	< .001	-0.74	.36	.013	47
3. Empathy vs. emotion recog.	-0.14 [-0.28, -0.01]	.041	-0.14	.43	< .001	196
4. Valence manip., between	-0.55 [-0.71, -0.40]	< .001	-0.51	.50	< .001	193
5. Valence manip., between	-0.43 [-0.59, -0.28]	< .001	-0.41	.26	< .001	206
6. Valence manip., within	-0.43 [-0.87, 0.01]	.057	-0.36	.50	< .001	49
7. Feel self/other, low arousal	-0.52 [-0.73, -0.30]	< .001	-0.50	.29	.005	91
8. Feel self/other, med. arousal	-0.79 [-1.03, -0.56]	< .001	-0.77	.39	< .001	87
9. Efficacy manipulation	-1.35 [-1.74, -0.97]	< .001	-1.45	.59	< .001	90
10. Efficacy manipulation	-1.43 [-1.80, -1.06]	< .001	-1.57	.42	< .001	93
11. Varying empathic demand	-0.20 [-0.34, -0.06]	.007	-0.20	.21	.004	192
Meta-analytic effect		< .001	**-0.54**	**.39**	< .001	1300

Pilot Study: Testing the Empathy Efficacy Manipulation

In a pilot study that preceded Studies 9 and 10 (N = 18), which used a similar efficacy manipulation and manipulation check, there was a marginally significant efficacy condition x efficacy type x time of measurement interaction, $F(1, 16) = 4.40$, $p = .052$, $\eta_p^2 = .22$. For empathy efficacy, there was an efficacy condition x time of measurement interaction, $F(1, 16) = 8.34$, $p < .011$, $\eta_p^2 = .34$, such that empathy efficacy significantly decreased over time in the low-efficacy condition ($M_{Time1} = 76.57$, $SD_{Time1} = 12.63$, $M_{Time2} = 64.00$, $SD_{Time2} = 17.57$, $F(1, 6) = 7.06$, $p = .038$, 95% CI = [-24.15, -1.00], $\eta_p^2 = .54$) and descriptively increased over time in the high-efficacy condition ($M_{Time1} = 69.91$, $SD_{Time1} = 26.19$, $M_{Time2} = 76.18$, $SD_{Time2} = 21.66$, $F(1, 10) = 2.19$, $p = .170$, 95% CI = [-3.17, 15.72], $\eta_p^2 = .18$), with the latter pattern non-significant likely due to low statistical power. By contrast, for emotion efficacy there was not an efficacy condition x time of measurement interaction, $F(1, 16) = 1.38$, $p = .257$, $\eta_p^2 = .08$, as emotion efficacy did not change over time in either condition (ps > .340). These pilot results led us to believe that this would be an effective and specific manipulation in the full samples.

Studies 9-10: Analyses of the Manipulation Checks for Empathy Efficacy

To examine the influence of the efficacy manipulation on empathy efficacy, we examined efficacy ratings for empathy and emotion self-awareness, which were measured both before and after the efficacy manipulation. We conducted a repeated measures ANOVA with time of measurement (pre- and post-manipulation) and efficacy type (empathy, emotion self-awareness) as within-subjects factors and efficacy condition as a between-subjects factor. If our manipulation influenced efficacy at empathy in particular, then effects should be stronger for empathy efficacy than emotion efficacy. As expected, in both Studies 9 and 10 there was an efficacy condition x efficacy type x time of measurement interaction (Study 9: $F(1, 88) = 13.39$, $p < .001$, $\eta_p^2 = .13$; Study 10: $F(1, 91) = 22.62$, $p < .001$, $\eta_p^2 = .20$). For empathy efficacy, there was an efficacy condition x time of measurement interaction (Study 9: $F(1, 88) = 25.85$, $p < .001$, $\eta_p^2 = .23$; Study 10: $F(1, 91) = 25.02$, $p < .001$, $\eta_p^2 = .22$), such that empathy efficacy decreased over time in the low-efficacy condition and increased over time in the high-efficacy condition. By contrast, for emotion efficacy, there was an efficacy condition x time of

measurement interaction in Study 9 ($F(1, 88) = 10.22$, $p = .002$, $\eta_p^2 = .10$), but not in Study 10 ($F(1, 91) = .01$, $p = .915$, $\eta_p^2 = .00$). See Supplemental Table S4 below. These results suggest that the efficacy manipulation is targeting empathy efficacy in particular.

Table S4. Efficacy Manipulation Checks, by Efficacy Condition and Time, Studies 9-10.

	Efficacy Time 1 *M* (*SD*)	Efficacy Time 2 *M* (*SD*)	*F*	*p*	95% CI	Hedges' *g*
Study 9						
Empathy, Low Efficacy	76.15 (16.62)	65.63 (23.02)	13.02	.001	[-16.40, -4.62]	-0.50
Empathy, High Efficacy	66.67 (22.22)	72.88 (20.11)	12.39	.001	[2.66, 9.75]	0.28
Emotion, Low Efficacy	88.10 (11.95)	85.05 (14.06)	5.61	.023	[-5.65, -.45]	-0.22
Emotion, High Efficacy	81.84 (19.32)	83.84 (16.71)	4.32	.043	[.07, 3.94]	0.10
Study 10						
Empathy, Low Efficacy	65.34 (23.45)	57.34 (24.85)	13.35	.001	[-12.42, -3.58]	-0.32
Empathy, High Efficacy	62.41 (22.84)	68.90 (21.02)	11.47	.001	[2.64, 10.34]	0.29
Emotion, Low Efficacy	79.00 (21.11)	79.82 (22.02)	1.39	.245	[-2.22, .58]	0.04
Emotion, High Efficacy	81.78 (14.83)	82.78 (15.69)	0.45	.504	[-1.99, 3.99]	0.06

Next, we examined the NASA Task Load index responses that occurred before and after the Empathy Selection Task test trials. These measures provided an additional opportunity to test the effectiveness of the efficacy manipulation on perceived cognitive costs of empathy. For each deck, we conducted repeated measures ANOVAs with time of measurement (pre-task, post-task) and deck type (empathy, objective) as within-subjects factors and efficacy condition as a between-subjects factor. For ease of presentation, we report the 2-way interactions averaging across time of measurement. As expected, efficacy condition moderated the influence of deck type on ratings of effort (Study 9: $F(1, 88) = 21.18$, $p < .001$, $\eta_p^2 = .19$; Study 10: $F(1, 91) = 58.67$, $p < .001$, $\eta_p^2 = .39$), aversion (Study 9: $F(1, 88) = 41.90$, $p < .001$, $\eta_p^2 = .32$; Study 10: $F(1, 91) = 73.14$, $p < .001$, $\eta_p^2 = .45$), and efficacy (Study 9: $F(1, 88) = 74.12$, $p < .001$, $\eta_p^2 = .46$; Study 10: $F(1, 91) = 82.89$, $p < .001$, $\eta_p^2 = .48$). Within the low-efficacy conditions, the empathy deck (vs. objective deck) was rated as higher on effort and aversion, and lower in efficacy, as in other studies. By contrast, within the high-efficacy conditions, the empathy (vs. objective) deck was not rated as different on effort, and was rated as lower in aversion and higher in efficacy. Supplemental Table S5 (below) displays descriptive and inferential statistics for these comparisons. Thus, the efficacy manipulation either minimized or reversed the typical cognitive costs associated with empathy (vs. objectivity). These findings internally replicate the effect of efficacy condition on the empathy efficacy manipulation check.

Table S5. NASA Task Load Index Ratings, by Efficacy Condition and Deck Type, Studies 9-10.

	Empathy Deck	Objective Deck	*F*	*p*	95% CI	Hedges' *g*
Study 9						
Effort, Low Efficacy	3.69 (1.03)	2.47 (1.13)	44.03	< .001	[0.85, 1.59]	1.11
Effort, High Efficacy	2.81 (0.94)	2.65 (0.96)	1.30	.259	[-0.12, 0.45]	0.17
Aversion, Low Efficacy	2.72 (1.27)	1.50 (0.64)	31.06	< .001	[0.78, 1.66]	1.19
Aversion, High Efficacy	1.71 (0.81)	2.18 (1.06)	9.39	.004	[-0.78, -0.16]	-0.48
Efficacy, Low Efficacy	2.87 (1.07)	4.22 (0.72)	49.57	< .001	[-1.74, -0.97]	-1.45
Efficacy, High Efficacy	4.16 (0.70)	3.51 (0.81)	21.76	< .001	[0.37, .0.94]	0.85
Study 10						
Effort, Low Efficacy	3.89 (0.72)	2.12 (1.07)	103.55	< .001	[1.42, 2.12]	1.89
Effort, High Efficacy	3.00 (1.06)	2.84 (1.06)	1.77	.190	[-0.08, 0.41]	0.15
Aversion, Low Efficacy	2.88 (1.09)	1.55 (0.83)	69.00	< .001	[1.01, 1.65]	1.33
Aversion, High Efficacy	1.92 (0.75)	2.27 (1.00)	8.68	.005	[-0.58, -0.11]	-0.38
Efficacy, Low Efficacy	2.86 (0.84)	4.30 (0.95)	60.33	< .001	[-1.80, -1.06]	-1.57
Efficacy, High Efficacy	4.22 (0.71)	3.62 (0.86)	20.80	< .001	[0.34, 0.87]	0.75

Study 11: Reward and Valuation Analyses. Novel to this study, we examined deck differences in emotional and social reward (averaged together) and value. Participants did not find the high-demand deck more rewarding than the low-demand deck (M_{high} = 2.29, SD_{high} = 1.25, M_{low} = 2.27, SD_{low} = 1.29, $F(1, 191) = .15$, $p = .696$), nor did they find it more valuable (M_{high} = 2.72, SD_{high} = 1.37, M_{low} = 2.76, SD_{low} = 1.25, $F(1, 191) = .26$, $p = .613$). However, participants were more likely to choose the high-demand deck to the degree they found it more valuable, $r = .20$, $p = .006$, with no association for felt reward, $r = -.01$, $p = .855$.

Table S6. Time Course Effects in Empathy Selection Task, Studies 1-11

	B	SE	*t*	*p*	Exp(B)	95% CI Exp (B)
Study 1	-.04	.01	-3.22	.001	0.96	0.94, 0.98
Study 2	-.08	.02	-4.46	< .001	0.93	0.89, 0.96
Study 3	-.01	.01	-1.66	.097	0.99	0.98, 1.00
Study 4	-.03	.01	-3.64	< .001	0.98	0.96, 0.99
Study 5	-.04	.01	-7.27	< .001	0.96	0.95, 0.97
Study 6	-.03	.02	-1.54	.124	0.97	0.93, 1.01
Study 7	-.02	.01	-4.66	< .001	0.98	0.97, 0.99
Study 8	-.02	.01	-3.77	< .001	0.98	0.97, 0.99
Study 9	.00	.01	0.28	.780	1.00	0.98, 1.03
Study 10	.02	.01	1.34	.179	1.02	0.99, 1.04
Study 11	-.03	.01	-4.41	< .001	0.97	0.96, 0.99
Meta-analytic	**-.02**	.01		< .001	0.98	0.97, 0.99

Note on Tables S6-S8. Results are from generalized linear mixed models in SPSS GENLINMIXED with centered time variable (ranging from 0-39 in Studies 1-8 and 11, and 0-23 in Studies 9-10), a random intercept and random slope for time, and an autoregressive covariance parameter. The high-efficacy conditions of Studies 9 and 10 are excluded because empathy avoidance was not predicted to occur in these conditions. The bottom row indicates results from random-effects meta-analysis.

Table S7. Moderation of Time Course Effects by NASA Effort Ratings, Studies 1-11.

	B	SE	*t*	*p*	Exp(B)	95% CI Exp (B)
Study 1	-.02	.01	-1.95	.052	0.98	0.96, 1.00
Study 2	-.04	.02	-2.02	.043	0.97	0.93, 1.00
Study 3	-.03	.01	-3.44	.001	0.97	0.95, 0.99
Study 4	-.02	.01	-3.25	.001	0.98	0.97, 0.99
Study 5	-.02	.01	-3.28	.001	0.99	0.98, 0.99
Study 6	-.00	.02	-0.06	.951	1.00	0.97, 1.03
Study 7	-.00	.01	-0.31	.760	1.00	0.99, 1.01
Study 8	-.01	.01	-1.98	.048	0.99	0.97, 1.00
Study 9	-.01	.01	-1.27	.203	0.99	0.97, 1.01
Study 10	.01	.01	1.02	.310	1.01	0.99, 1.03
Study 11	-.02	.01	-1.59	.112	0.98	0.97, 1.00
Meta-analytic	**-.01**	.00		< .001	0.99	0.98, 0.99

Table S8. Moderation of Time Course Effects by NASA Aversion Ratings, Studies 1-11.

	B	SE	*t*	*p*	Exp(B)	95% CI Exp (B)
Study 1	-.01	.01	-0.81	.420	0.99	0.98, 1.01
Study 2	-.03	.02	-1.63	.104	0.98	0.95, 1.01
Study 3	-.04	.01	-4.04	< .001	0.96	0.95, 0.98
Study 4	-.03	.01	-3.76	< .001	0.98	0.96, 0.99
Study 5	-.02	.01	-2.98	.003	0.99	0.98, 1.00
Study 6	-.02	.02	-0.97	.331	0.99	0.96, 1.02
Study 7	-.00	.01	-0.39	.699	1.00	0.99, 1.01
Study 8	-.01	.01	-1.12	.264	1.00	0.99, 1.00
Study 9	-.02	.01	-2.31	.021	0.99	0.97, 1.00
Study 10	.00	.01	0.05	.960	1.00	0.98, 1.02
Study 11	-.01	.00	-2.32	.021	0.99	0.98, 1.00
Meta-analytic	**-.01**	.00		< .001	0.99	0.98, 0.99

Table S9. Moderation of Time Course Effects by NASA Efficacy Ratings, Studies 1-11.

	B	SE	*t*	*p*	Exp(B)	95% CI Exp (B)
Study 1	.01	.01	0.78	.438	1.01	0.99, 1.03
Study 2	.02	.02	1.35	.179	1.02	0.99, 1.06
Study 3	.04	.01	4.55	< .001	1.04	1.02, 1.06
Study 4	.03	.01	4.53	< .001	1.03	1.02, 1.05
Study 5	.02	.01	3.34	.001	1.02	1.01, 1.03
Study 6	.04	.01	3.90	< .001	1.04	1.02, 1.06
Study 7	.01	.01	2.02	.043	1.01	1.00, 1.02
Study 8	.01	.01	1.97	.049	1.01	1.00, 1.02
Study 9	.03	.01	2.33	.020	1.03	1.00, 1.05
Study 10	-.00	.01	-0.30	.761	1.00	0.98, 1.01
Study 11	.01	.01	2.20	.028	1.01	1.00, 1.03
Meta-analytic	**.02**	.00		< .001	1.02	1.01, 1.03

III. SUPPLEMENTAL METHOD FOR SECONDARY STUDIES

In addition to the 11 studies reported in the main text, we conducted an additional 11 studies testing different variations of the Empathy Selection Task. Below, we report the method for these additional studies, and then report supplementary meta-analyses of empathy choice and cognitive costs including the full set of studies.

Supplemental Studies S1-S3

Participants. In Study S1, we enrolled 190 MTurk participants. From this sample, 81 participants dropped out of the survey before finishing the Empathy Selection Task. We excluded 7 participants who skipped at least one response on the Empathy Selection Task, leaving a final sample of 102 participants (61 female, 38 male, 1 other, 2 unreported, M_{age} = 40.85, SD_{age} = 14.75). In Study S2, we enrolled 186 MTurk participants. From this initial sample, 75 participants dropped out before finishing the Empathy Selection Task. We excluded 4 participants who skipped at least one response on the Empathy Selection Task, leaving a final sample of 107 participants (60 female, 46 male, 1 unreported, M_{age} = 37.59, SD_{age} = 11.89). In Study S3, we enrolled 134 MTurk participants. From this initial sample, 57 participants dropped out before finishing the Empathy Selection Task. We excluded 3 participants who skipped at least one response on the Empathy Selection Task, leaving a final sample of 74 participants (53 female, 21 male, M_{age} = 39.39, SD_{age} = 14.14).

Empathy Selection Task. Pre-task instructions were nearly identical to Study 1. In Study S1, the pre-task instruction "you will then see an image of a person" read "a person's hand", and on one trial the trial-level instruction said to "look at the hand in the picture" instead of "look at the person in the picture". In Studies S2-S7, there were four trials with the same trial-level typo. Participants completed 25 trials in Study S1 and 40 trials in Studies S2-S3. In Study S3, decks were unlabeled as "DECK 1" and "DECK 2" rather than as "FEEL" and "DESCRIBE". For that study, in the pre-task instructions, the beginning of the second and third paragraphs removed reference to deck labels, instead saying "On some trials…" and "On other trials…" In Supplemental Studies S2-S4, there was a programming error on one of the empathy trials: although the instructions encouraged participants to feel empathy, it requested an age/gender response. This typo was fixed in subsequent studies. Because the pattern of avoidance behavior emerges across 40 trials, we do not believe that this rare typo had a significant influence on choice behavior in the Empathy Selection Task.

Post-Task Questionnaire. After the Empathy Selection Task, participants provided open-ended responses to the same four open-ended questions as in the main studies. Additionally, they responded to two questions: "For some participants, one of the two decks had a tendency to contain more 'feeling' trials while the other deck had less 'feeling' trials. Did it seem like this was the case for you? If so, which deck tended to have more 'feeling' trials (red deck or blue deck)?" "If you answered yes to the previous question (indicating that one of the decks seemed to have more 'feeling' trials), was this something you became EXPLICITLY are of DURING THE EXPERIMENT, or something that you realized only in retrospect?"

NASA Task Load Index. This was identical to the main studies. In Study S3, which used unlabeled decks, decks were referred to by color and position (e.g., "the red deck (the one on the left)" and "the blue deck (the one on the right)") rather than by name.

Additional Measures. In all three studies, participants completed the Interpersonal Reactivity Index. In Study S3, participants additionally completed the Empathy Quotient Short-Form and the Fear of Compassion for Others Scale. Individual differences were measured after the Empathy Selection Task in Study S1 and before the Empathy Selection Task in Studies S2-S3. Demographics were assessed at the end, as in the main studies.

Supplemental Study S4 (Probabilistic Decks)

Participants. We enrolled 100 MTurk participants. From this initial sample, 43 participants dropped out before finishing the Empathy Selection Task. We excluded 3 participants who skipped at least one response on the Empathy Selection Task, leaving a final sample of 54 participants (33 female, 21 male, M_{age} = 38.04, SD_{age} = 12.84).

Empathy Selection Task. This was nearly identical to Study S3 except that a note was added at the end of the pre-task instructions to go at one's own pace. Whereas in all other main and supplemental studies the right deck contained 100% empathy trials and the left deck contained 100% objective trials, in this study the right deck contained 70% empathy trials and 30% objective trials, whereas the left deck contained 30% empathy trials and 70% objective trials. We expected that although there would be an empathy avoidance effect, that it would not be as strong as previous effects given the lack of uniform deck composition, and because participants may have taken more time to learn deck composition across trials.

Post-Task Questionnaire This was identical to previous studies.

NASA Task Load Index. This was identical to Study S3.

Additional Measures. Prior to the Empathy Selection Task, participants completed the Interpersonal Reactivity Index, Empathy-Quotient Short-Form, and Fear of Compassion for Others Scale. Demographics were assessed as in previous studies.

Supplemental Study S5 (Labels, One-Word Response)

Participants. We enrolled 258 MTurk participants. From this initial sample, 48 participants dropped out before finishing the Empathy Selection Task. We excluded 5 participants who skipped at least one response on the Empathy Selection Task, and 5 participants who had previously completed an earlier study in this sequence, leaving a final sample of 200 participants (127 female, 71 male, 1 other, 1 unreported, M_{age} = 37.05, SD_{age} = 11.14).

Empathy Selection Task. The task was identical to the task used in Study S2, except for the inclusion of the note about pacing that was incorporated in Study S4. We modified the instructions to test the alternative explanation that people are avoiding having to verbalize a full sentence about the internal experiences and feelings of others. The sentence prompt was

included in the standard version of the Empathy Selection Task in order to increase immersion and engagement with targets of empathy, and so make the psychological costs of empathy more salient to participants. In the revised task, participants were instructed to make a single-word, rather than full-sentence, response on each trial. We expected to observe empathy avoidance, but that the effect would be weaker due to reduced immersion on each trial. Participants read the following pre-task instructions:

> In this task, you will complete a series of trials. On each trial, you will see two decks of cards: the deck on the left will always be labeled "DESCRIBE" and the deck on the right will always be labeled "FEEL". You should choose between these decks. Once you choose a deck, will then see an image of a person. Depending on which deck you have chosen, you will be given one of two possible sets of instructions.
>
> If you choose from the deck labeled "DESCRIBE", you will be told to be objective and focus on the external features and appearances of the person in the image. When completing this kind of trial, try to be as objective as possible. To be objective, do not let yourself get caught up in imagining what this person feels. On these trials, use ONE WORD to describe the age of the person. Please limit your response to a single word. For example, you could write "two" to indicate that you think the person in the image is two years old; or you could write "fifteen" to indicate that you think the person in the image is fifteen years old.
>
> If you choose from the deck labeled "FEEL", you will be told to have empathy and focus on the internal feelings and experiences of the person in the image. When completing this kind of trial, try to feel as much empathy as possible. To be empathic, let yourself get caught up in imagining what this person feels. On these trials, use ONE WORD to describe the feelings of the person. Please limit your response to a single word. For example, you could write "pain" to indicate that you think the person in the image is feeling pain; or you could write "pleasure" to indicate that you think the person in the image is feeling pleasure.
>
> You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. Overall, this task will take the same amount of time regardless of which deck you choose. You should go at your own pace.

On "DESCRIBE" trials, the last line of the trial-level instructions was replaced with "Please write one word describing the age of this person." On "FEEL" trials, the last line of the instructions was replaced with "Please write one word describing the feelings of this person."

Post-Task Questionnaire. The post-task questionnaire was identical to previous studies, except that the two questions about awareness of decks were removed (here and in all subsequent supplementary studies)**.**

NASA Task Load Index. This was identical to previous studies, except that the objective deck was referred to "the Describe deck (the one on the left)" and the empathy deck was referred to as "the Feel deck (the one on the right)".

Additional Measures. Participants completed the IRI and Empathy-Quotient Short-Form prior to the Empathy Selection Task. Demographics were assessed as in previous studies.

Supplemental Study S6 (Changing Opportunity Costs)

Participants. We enrolled 108 MTurk participants. From this initial sample, 53 participants dropped out before finishing the Empathy Selection Task. We excluded 1 participant who had previously completed an earlier study in this sequence, leaving a final sample of 54 participants (37 female, 16 male, 1 other, M_{age} = 34.37, SD_{age} = 10.71).

Empathy Selection Task. In all studies using the Empathy Selection Task, participants were given a choice between experiencing and avoiding empathy. Yet relative costs of empathy may shift depending on the context of choice. When the contrasting non-empathy choice is comparably effortful, opportunity costs should change and people should be less likely to avoid empathy (Kurzban, Duckworth, Kable, & Myers, 2013). Opportunity costs reflect the value that is lost from pursuing a current course of action when an alternative course of action is available (Kurzban et al., 2013). When the opportunity cost of a given course of action is high—as it would be when effortful empathy is contrasted against a less effortful alternative such as objectivity—people may prioritize competing courses of action, engaging in empathy avoidance. When the opportunity cost of a given course of action is low—as it would be if effortful empathy is contrasted against an equally effortful alternative—people may be less likely to prioritize competing courses of action, reducing empathy avoidance. In Study S6, we adapted the Empathy Selection Task so that objective deck choices involved instructions to write one sentence about the health and hygiene of the target. Writing about health may be more effortful and aversive because it requires making inferences that are not as apparent as identifying gender and age. The task was identical to Study S3, except that pre-task instructions were changed for objective trials: "On these trials, use one sentence to describe the physical health and hygiene of the person." When participants selected the Describe deck during the task, the last sentence of the trial-level instructions was altered: "… Please write one sentence describing the health and hygiene of this person." We expected that in this study, empathy avoidance would be reduced because empathy would no longer seem to be the most effortful option.

Post-Task Questionnaire This was identical to Study S5.

NASA Task Load Index. This was identical to Study S5.

Additional Measures. Demographics were assessed as in previous studies.

Supplemental Study S7 (Three-Word Response Pilot)

Participants. This study was a pilot for the Empathy Selection Task variant that was used with the efficacy manipulations in Studies 9-10 in the main text. We enrolled 29 MTurk

participants. From this initial sample, 8 participants dropped out before finishing the Empathy Selection Task. We excluded 3 participants who had completed an earlier study in this sequence, leaving a final sample of 18 participants (7 female, 11 male, M_{age} = 33.89, SD_{age} = 12.86).

Empathy Selection Task. The task instructions were nearly identical to those used in Studies 9 and 10 in the main text. The main difference between this and other versions of the Empathy Selection Task is that responses were changed to be 3 keywords rather than a full sentence. The change was implemented to simplify the response so that the same amount of information would be provided across decks, and to make accuracy feedback seem plausible to participants for the efficacy manipulations. Pre-task instructions read:

> In this task, you will complete a series of trials. On each trial, you will see two decks of cards. You should choose between these decks. Once you choose a deck, you will then see an image of a person. Depending on which deck you have chosen, you will be given one of two possible sets of instructions.
>
> On the DESCRIBE trials, you will be told to be objective and focus on the external features and appearances of the person in the image. When completing this kind of trial, try to be as objective as possible. To be objective, do not let yourself get caught up in imagining what this person feels. On these trials, please provide three keywords to describe the physical appearance of the person, as if you were describing them to a sketch artist. (Example: "old, white, woman" or "young, black, man")
>
> On the FEEL trials, you will be told to have empathy and focus on the internal feelings and experiences of the person in the image. When completing this kind of trial, try to feel as much empathy as possible. To be empathic, let yourself get caught up in imagining what this person feels. On these trials, please provide three keywords to describe the feelings and experiences of the person. (Example: "sad, hurt, confused" or "happy, pleased, interested")
>
> You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. Overall, this task will take the same amount of time regardless of which deck you choose.

The comprehension check, trial-level instructions and Empathy Selection Task parameters were similar to Studies 9-10 in the main text.

Post-Task Questionnaire. This was identical to Study S5.

NASA Task Load Index. This was identical to Study S5.

Additional Measures. Demographics were assessed as in previous studies.

Supplemental Study S8 (Valence Manipulation 4)

Participants. We enrolled 265 MTurk participants. From this initial sample, 63 participants dropped out before finishing the Empathy Selection Task. Of these participants who dropped out, 14 were in the positive-valence/empathy-on-right condition, 18 were in the negative-valence/empathy-on-right condition, 13 were in the negative-valence/empathy-on-left condition, and 18 were in the positive-valence/empathy-on-left condition. We excluded 1 participant who skipped at least one response on the Empathy Selection Task, and 5 participants who had previously completed an earlier study in this sequence, leaving a final sample of 196 participants (94 female, 101 male, 1 other, M_{age} = 34.04, SD_{age} = 9.44).

Empathy Selection Task. The task was nearly identical to the task used in Supplemental Study S7 and in Studies 9-10 in the main text, using the three-keywords response prompt. In this study, decks were both colored blue and were unlabeled ("DECK 1" and "DECK 2"), with. Deck position was counterbalanced, such the empathy deck was manipulated to be either on the left or right. These changes were implemented to rule out color and deck position preferences as alternative explanations of empathy avoidance. There was a between-subjects manipulation of target valence (negative, positive), focusing on happiness vs. sadness. Target images were White female/male actors from the NimStim Database (Tottenham et al., 2009), posing either happy or sad expressions (White female exemplars: 01F, 02F, 03F, 05F, 06F, 07F, 08F, 09F, 10F; White male exemplars: 20M, 21M, 22M, 23M, 24M, 25M, 26M, 28M, 29M, 30M, 31M, 32M, 33M, 34M, 36M, 37M). Unlike in Studies 4-5 in the main text, in Study S8 the positive condition did not contain 100% positive exemplars and the negative condition did not contain 100% negative exemplars. Instead, 16 of 25 trials contained positive targets in the positive condition, with the other 9 trials containing negative targets; and vice versa in the negative condition. To emphasize the dominant valence within each condition, trials were presented in a pseudo-random order: within the positive condition, the first five trials always went positive, negative, positive, positive, positive, and were randomized thereafter; and vice versa in the negative condition. The task instructions were nearly identical to those in Study S7 and Studies 9-10. We altered the second sentences of the first and second paragraph, so that instead of instructing participants to be as objective as possible or feel as much empathy as possible, they were instructed to "try to be objective" and to "try to feel empathy." Additionally, at the end of the third paragraph, the last sentence was amended to read: "It is okay to use the same keyword multiple times, just make sure you are describing the feelings and experiences of the person in the image (e.g., mood, emotion, etc.)" References to deck labels in the instructions were removed, as in other studies using unlabeled decks. Participants completed the same comprehension check as in Study S7 and Studies 9-10, with the same response options, however the questions were revised given the lack of deck labels: "Which of the following is an appropriate response on trials where you are told to be objective?" "Which of the following is an appropriate response on trials where you are told to be empathic?" Trial-level instructions were identical to those used in the Study S7 and Studies 9-10. Participants completed 25 trials.

Post-Task Questionnaire. This was identical to Study S5.

NASA Task Load Index. This was identical to Study S5. There was a typo, such that in the empathy-on-left conditions, the empathy deck was referred to as "the FEEL deck" and the objective deck as "the DESCRIBE deck", instead of "Deck 1" and "Deck 2" as in the empathy-

on-right conditions. Because deck position did not moderate the influence of deck type on NASA cost measures (*p*s > .280), we do not believe this typo mattered.

Donation. Participants were then asked: "How much of your MTurk earnings today would you be willing to donate to the American College Health Foundation, an organization devoted to improving the psychological health of college students?" Participants could respond in $0.01 increments, with scale points anchored at $0.00 and $2.00 and marked every $0.25.

Empathy Discounting Paradigm. Participants completed an Empathy Discounting Paradigm nearly identical to the version in Study 5 in the main text. The only difference is that a typo in the instructions was fixed, so that instead of referring to crying children the instructions simply said "In all cases, the persons shown will be similar to those you saw earlier in the experiment." Deck labels were retained in order to simplify the procedure and ensure that participants knew what psychological tasks each price amount was being attached to; as in Study 5, the pre-task instructions reiterated what was expected on the FEEL and DESCRIBE decks.

Additional Measures. Participants completed the 18-item Short Need for Cognition Scale (Cacioppo, Petty, & Kao, 1984; e.g., "I find satisfaction in deliberating hard and for long hours"). Demographics were assessed as in previous studies.

Supplemental Study S9 (Counterbalanced Deck Positions)

Participants. We enrolled 268 MTurk participants. From this initial sample, 64 participants dropped out before finishing the Empathy Selection Task. Of those who dropped out, 29 were in the empathy-on-right condition and 35 were in the empathy-on-left condition. We excluded 4 participants who had completed an earlier study, leaving a final sample of 200 participants (105 female, 93 male, 1 other, 1 unreported, M_{age} = 34.73, SD_{age} = 11.12).

Empathy Selection Task. The task was very similar to that used in Study S8, using the three-keywords response prompt. However, instead of making both decks blue, a feature which was unique to Study 15, we reverted to the standard depiction with the left deck being red and the right deck being blue. We retained unlabeled decks, but the left deck was labeled "Deck A" and the right deck was labeled "Deck B". As in Study S8, there was a between-subjects manipulation of deck position, such that half of participants had the empathy deck on the right and the other half had the empathy deck on the left. The pre-task instructions were nearly identical to those used in Study S8, except that the third sentence of the second paragraph was replaced with: "To be objective, try to focus on the person's external appearances", and the third sentence of the third paragraph was replaced with: "To be empathic, try to share the person's internal experiences." This phrasing replaced prior language that instructed participants to get caught up in imagining what the person feels; we made this change in order to remove any language that might seem to encourage perspective-taking instead of experience sharing. Participants completed the same comprehension check as in Study S8. Trial-level instructions were nearly identical to those in the Study S7-S8 and Studies 9-10. On empathy trials, instructions were altered to more closely match experience sharing instructions from published work (Klimecki et al., 2013): "Look at the person in the picture, and **try to feel what this person feels**. Empathically share in the internal experiences and feelings of this person." Target images

were Black and White male and female exemplars from the Chicago Face Database displaying anger, identical to the stimuli in Studies 4-5 in the main text. Participants completed 40 trials.

Post-Task Questionnaire. This was identical to Study S5.

NASA Task Load Index. This was identical to Study S5, except that the decks were referred to as "Deck A" and "Deck B".

Social Norms. Participants were told: "Please answer the following questions, as honestly as possible, based on your own personal opinion and experiences." First, they were asked about descriptive norms of choice in the Empathy Selection Task (scale from 0-100): "What percentage of people tend to choose the empathy deck in the task you just completed?" "What percentage of people tend to choose the objective deck in the task you just completed?" Second, they were asked about injunctive norms about empathy and objectivity (scale from 0-100): "What percentage of people think that empathy is a good thing?" "What percentage of people think that objectivity is a good thing?" Finally, participants were asked: "According to your own personal beliefs, do you think that it is more desirable to be empathic or to be objective" (1=*objectivity is much more desirable*, 2=*objectivity is more desirable*, 3=*they are equally desirable*, 4=*empathy is more desirable*, 5=*empathy is much more desirable*).

Additional Measures. Demographics were assessed as in previous studies.

Supplemental Study S10 (Valence Manipulation 5)

Participants. We enrolled 262 MTurk participants. From this initial sample, 58 participants dropped out before finishing the Empathy Selection Task. Of these participants who dropped out, 14 were in the positive-valence/empathy-on-right condition, 12 were in the negative-valence/empathy-on-right condition, 20 were in the positive-valence/empathy-on-left condition, and 12 were in the negative-valence/empathy-on-left condition. We excluded 2 participants who had previously completed an earlier study in this sequence, leaving a final sample of 202 participants (111 female, 89 male, 2 unreported, M_{age} = 38.14, SD_{age} = 12.46).

Empathy Selection Task. The task was nearly identical to Study S9. Decks were unlabeled (as "Deck A" and "Deck B") and empathy deck position was counterbalanced. Like Study S8, there was also a between-subjects manipulation of target valence. Target images were actors from the NimStim Database, posing either happy or sad expressions (White female exemplars: 01F, 02F, 03F, 05F, 06F, 07F, 08F, 09F, 10F; White male exemplars: 20M, 21M, 22M, 23M, 24M, 25M, 26M, 28M, 29M, 30M, 31M, 32M, 33M, 34M, 36M, 37M; Black female exemplars: 11F, 12F, 13F, 14F; Black male exemplars: 38M, 39M, 40M, 41M, 42M, 43M; Asian female exemplars: 15F, 16F, 17F, 18F, 19F). Unlike Study S8, the manipulation was structured like the other two valence manipulation studies (Studies 4 and 5 in main text), with all trials within each valence condition containing the target valence (i.e., all trials in the positive-valence condition displayed happy expressions), and trials were fully randomized. After the pre-task instructions, participants completed the same comprehension check as in Studies S8 and S9.

Post-Task Questionnaire. This was identical to Study S5.

NASA Task Load Index. This was identical to Study S9.

Social Norms. Participants completed the norms questions as in Study S9.

Additional Measures. Participants completed the Interpersonal Reactivity Index. Demographics were assessed as in previous studies.

Supplemental Study S11 (Vignette Stimuli)

Participants. We enrolled 263 MTurk participants. From this initial sample, 60 participants dropped out before finishing the Empathy Selection Task. We excluded 14 participants who had previously completed an earlier study in this sequence, leaving a final sample of 189 participants (67 female, 120 male, 2 unreported, M_{age} = 35.98, SD_{age} = 11.38).

Empathy Selection Task. Participants completed a modified variant of the Empathy Selection Task, receiving the following pre-task instructions:

> In this task, you will complete a series of trials. On each trial, you will see the name of a person, and a hypothetical event that could happen to that person. Try to visually imagine this event. You will also see two decks of cards: one deck will be labeled "DESCRIBE" and the other deck will be labeled "FEEL". You should choose between these decks. Depending on which deck you have chosen, you will be given one of two possible sets of instructions.
>
> If you choose from the red deck labeled "DESCRIBE", you will be told to be objectively detached toward the person. When completing this kind of trial, try to remain as objective as possible toward this person. To be objective, try to focus on external details of the event happening to the person. Once you have become fully objective, you can advance to the next trial.
>
> If you choose from the blue deck labeled "FEEL", you will be told to have empathy for the person. When completing this kind of trial, try to feel as much empathy as possible for this person. To be empathic, try to share in the person's feelings and experiences. Once you have fully generated feelings of empathy, you can advance to the next trial.
>
> You are free to choose from either deck on any trial, and should feel free to move from one deck to the other whenever you choose. If one deck begins to seem preferable, feel free to choose that deck more often. Overall, this task will take the same amount of time regardless of which deck you choose. Note also that the decks will switch sides over the course of the task.

At the beginning of each trial of the task, participants were shown a pair of card decks. There was a red deck labeled “DESCRIBE” and a blue deck labeled “FEEL”. The positioning of each deck on left vs. right was randomly counterbalanced across trials. As in previous studies, there

was not a time limit on choice. Above the card decks, participants also saw one of a series of event descriptions (from Bruneau, Cikara, & Saxe, 2015). These included 16 negative events (e.g., "Diana had a stomach ache after her lunch", "Bill sat on a stray nail on a bench") and 16 positive events (e.g.,"Jane just missed being hit by a bus", "Andrew found $500 in the street") of both mild and extreme intensity. Once a choice was made, participants saw the event description again. If participants selected the DESCRIBE deck, they were instructed: "Think about this person, and **try to be objective toward them. Focus on external details of what is happening to this person.** Once you have become OBJECTIVE, press continue." If participants selected the FEEL deck, they were instructed: "Think about this person, and **try to feel what this person feels. Share the feelings and experiences of this person.** Once you have generated EMPATHY, press continue." A Qualtrics timer was incorporated so that participants could not advance until after 5 seconds had elapsed. After entering a written response, participants completed state ratings of arousal ("How calm/aroused do you currently feel?" from 1=*calm* to 9=*aroused*) and valence ("How negative/positive do you currently feel?" from 1=*negative* 9=*positive*) which were accompanied by self-assessment manikin images (Bradley & Lang, 1994). Participants completed 32 trials which were presented in randomized order. There were 30 participants who failed to provide one of the state affect responses on at least one trial.

Post-Task Questionnaire. This was identical to previous studies.

NASA Task Load Index. Participants completed identical questions from earlier. In addition to the questions about effort, aversiveness, and efficacy, participants also completed three new questions per deck: "How emotionally rewarding was this deck?" "How socially rewarding was this deck?" "How valuable did you find this deck?"

Additional Measures. We assessed trait empathy on the IRI and Empathy Index, the Fear of Compassion for Others Scale, and demographics.

IV. SUPPLEMENTAL ANALYSES FOR SECONDARY STUDIES.

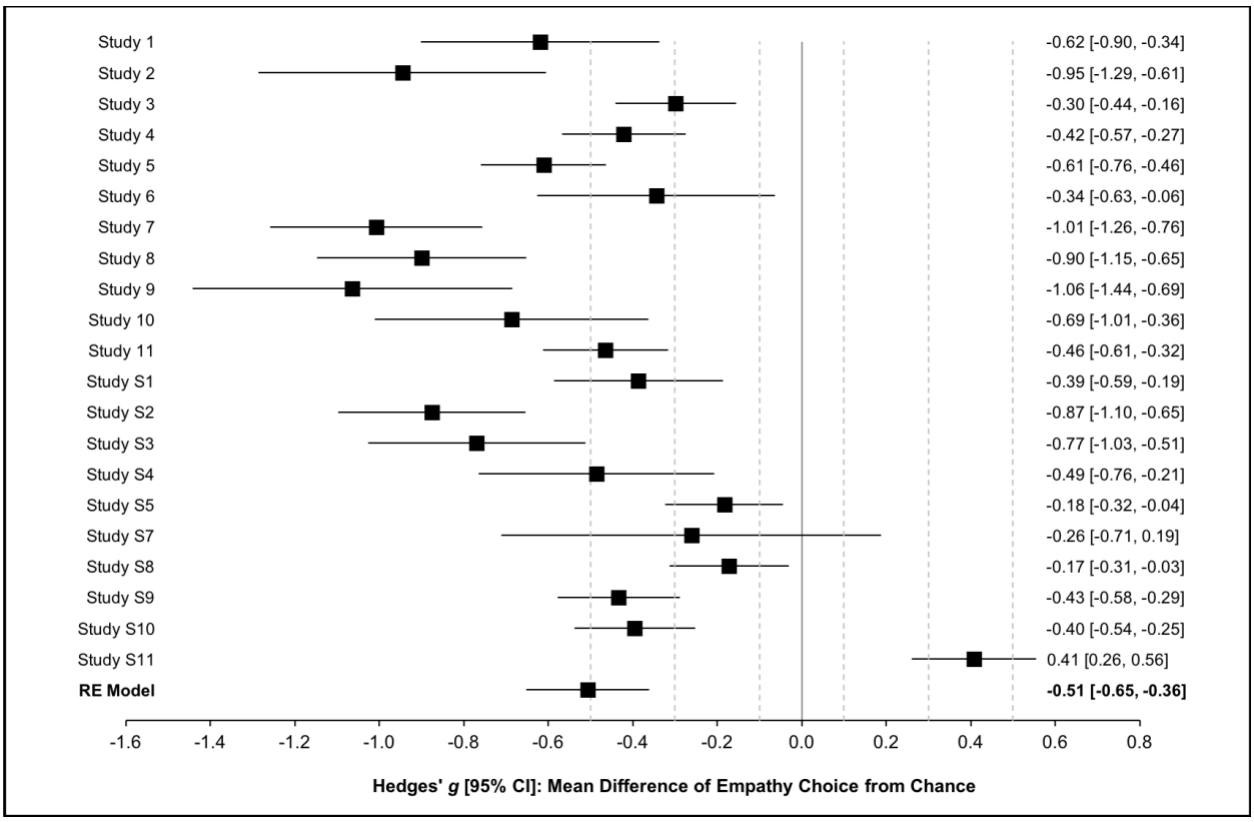
Lastly, we report meta-analyses of empathy choice across all studies, including those in the main text (Studies 1-11) and supplement (Supplemental Studies S1-S11). These analyses excluded Supplemental Study S6, because empathy avoidance was expected to be reduced when opportunity costs were altered. Table S6 depicts empathy choice across studies (N = 2,546). We examined mean difference of empathy choice from chance (50%, indicating no preference), with Hedges' g reflecting whether this mean difference deviated from zero, such that negative values indicate empathy avoidance. Using random-effects meta-analysis, the standardized mean difference of empathy choice (in Hedges' g) was -.51, 95% CI [-.65, -.36], Z = -6.80, p < .001, a large and robust empathy avoidance effect. Figure S1 displays the meta-analytic forest plot.

Table S10. Empathy choice in all studies (Studies 1-11, Supplemental Studies S1-S11).

Study	Emp. Choice *M* (*SD*)	95% CI M_{diff}	*t*	*N*	*p*	Hedges' *g*
1. Labeled decks	0.33 (0.27)	[-0.24, -0.10]	-4.70	56	< .001	-0.62
2. Non-labeled decks	0.26 (0.25)	[-0.31, -0.16]	-6.59	47	< .001	-0.95
3. Empathy vs. emotion recog.	0.41 (0.29)	[-0.13, -0.05]	-4.19	196	< .001	-0.30
4. Valence manip., between	0.38 (0.29)	[-0.17, -0.08]	-5.87	193	< .001	-0.42
5. Valence manip., between	0.34 (0.26)	[-0.19, -0.12]	-8.81	206	< .001	-0.61
6. Valence manip., within	0.38 (0.35)	[-0.22, -0.02]	-2.47	50	.017	-0.34
7. Feel self/other, low arousal	0.30 (0.19)	[-0.24, -0.16]	-9.69	91	< .001	-1.01
8. Feel self/other, med. arousal	0.34 (0.18)	[-0.20, -0.12]	-8.46	87	< .001	-0.90
9. Efficacy manipulation	0.22 (0.25)	[-0.36, -0.20]	-6.94	41	< .001	-1.06
10. Efficacy manipulation	0.33 (0.24)	[-0.24, -0.09]	-4.64	44	< .001	-0.69
11. Varying empathic demand	0.38 (0.26)	[-0.16, -0.08]	-6.48	193	< .001	-0.46
S1. Labeled decks	0.39 (0.29)	[-0.17, -0.06]	-3.94	102	< .001	-0.39
S2. Labeled decks	0.26 (0.27)	[-0.29, -0.18]	-9.11	107	< .001	-0.87
S3. Non-labeled decks	0.29 (0.27)	[-0.27, -0.15]	-6.69	74	< .001	-0.77
S4. Probabilistic decks	0.38 (0.24)	[-0.18, -0.05]	-3.62	54	.001	-0.49
S5. One-word response	0.45 (0.29)	[-0.09, -0.01]	-2.61	200	.010	-0.18
S7. Three-word response pilot	0.40 (0.36)	[-0.28, 0.08]	-1.16	18	.261	-0.26
S8. Valence manip., between	0.45 (0.27)	[-0.09, -0.01]	-2.41	196	.017	-0.17
S9. Counterbalanced deck pos.	0.37 (0.29)	[-0.17, -0.09]	-6.15	200	< .001	-0.43
S10. Valence manip., between	0.39 (0.28)	[-0.15, -0.07]	-5.64	202	< .001	-0.40
S11. Vignette stimuli	0.60 (0.24)	[0.06, 0.13]	5.63	189	< .001	0.41
Meta-analytic effect				2546	< .001	**-0.51**

Note. Studies 9 and 10 only include low-efficacy conditions because empathy avoidance was expected to be reduced in the high-efficacy conditions. Study S6 was not included because empathy avoidance was expected to be reduced when opportunity costs were changed.

Figure S1. Meta-analytic forest plot of empathy choice across all studies.



The role of opportunity costs in Study S6. Unlike previous studies, participants in Study S6 (which involved a harder non-empathy control deck) did not avoid empathy, choosing the empathy deck 55.51% of the time (SD = 27.04%), t = 1.50, p = .140, 95% CI [-.02, .13], Hedges' g = .20. As expected, changing the choice context to make the alternative to empathy less attractive and more effortful—i.e., describing the health of child refugees—reduced empathy avoidance. Unlike in previous studies, participants did not rate the empathy and objective decks as differing in effort (M_{Emp} = 3.77, SD_{Emp} = 1.06, M_{Obj} = 3.57, SD_{Obj} = .99, F(1, 53) = 2.73, p = .105, 95% CI [-.04, .43], η_p^2 = .05), or aversion (M_{Emp} = 3.44, SD_{Emp} = 1.33, M_{Obj} = 3.26, SD_{Obj} = 1.32, F(1, 53) = 1.81, p = .184, 95% CI [-.09, .46], η_p^2 = .03), and did not feel different levels of efficacy completing them (M_{Emp} = 3.67, SD_{Emp} = .99, M_{Obj} = 3.50, SD_{Obj} = 1.04, F(1, 53) = 3.12, p = .083, 95% CI [-.02, .36], η_p^2 = .06). By changing the choice context, the response options in the Empathy Selection Task were no longer seen as having different costs.

Target valence manipulations in Studies S8 and S10. In Studies S8 and S10, which contrasted sadness and happiness, participants showed non-significant avoidance of negative empathy (Study S8: $M_{EmpChoice}$ = 46.58%, $SD_{EmpChoice}$ = 27.57%, 95% CI M_{diff} = [-.09, .02], t = -1.21, p = .228, Hedges' g = -.12; Study S10: $M_{EmpChoice}$ = 45.28%, $SD_{EmpChoice}$ = 27.32%, 95% CI M_{diff} = [-.10, .01], t = -1.71, p = .090, Hedges' g = -.17) but avoidance of positive empathy (Study S8: $M_{EmpChoice}$ = 44.08%, $SD_{EmpChoice}$ = 26.97%, 95% CI M_{diff} = [-.11, -.01], t = -2.20, p = .030, Hedges' g = -.22; Study S10: $M_{EmpChoice}$ = 32.69%, $SD_{EmpChoice}$ = 27.78%, 95% CI M_{diff} = [-.23, -.12], t = -6.35, p < .001, Hedges' g = -.62). The non-significant effects for empathy

avoidance toward sad targets differ from earlier studies which did find avoidance of empathy for sad child refugees. The smaller avoidance effect in Study S8 may have been due to the fact that the sad condition contained majority (64%) sad trials but not 100% sad trials as in other studies; and in both studies, it may be that the sad expressions conveyed in the NimStim Database are less intense than those displayed by the child refugees. Avoidance rates did not differ by valence in Study S8, $F(1, 194) = .41$, $p = .521$, 95% CI [-.10, .05], $\eta_p^2 = .00$, but were stronger for positive empathy in Study S10, $F(1, 200) = 10.53$, $p = .001$, 95% CI [-.20, -.05], $\eta_p^2 = .05$.

Deck position manipulations in Studies S8-S10. Lastly, position of the empathy and objective decks on left vs. right was counterbalanced in Studies S8-S10; because Studies S9 and S10 use the standard color configuration, with the left deck being red and the right deck being blue, these studies also test whether color preferences explain the empathy choice effect. Deck position did not influence empathy choice (Study S8: $F(1, 194) = .25$, $p = .618$, 95% CI [-.10, .06], $\eta_p^2 = .00$; Study S9: $F(1, 198) = .00$, $p = .999$, 95% CI [-.08, .08], $\eta_p^2 = .00$; Study S10: $F(1, 200) = .573$, $p = .450$, 95% CI [-.11, .05], $\eta_p^2 = .00$), and effect sizes for empathy avoidance were comparable to other studies (see Figure S1 and Table S10), suggesting that deck position and color preferences do not explain the empathy avoidance effect observed across studies.

Empathy Discounting Paradigm in Study S8. Participants completed the Empathy Discounting Paradigm in Study S8, much as they did in Study 5 in the main text. This task, which was completed after the Empathy Selection Task, allowed us to financially quantify the comparative subjective costs assigned to the empathy deck in the Empathy Selection Task. The average indifference point across participants was \$1.72 (SD = \$0.49) in Study S8, and subjective cost of empathy was computed as \$2.00 minus the indifference point. Because the distribution was skewed (Study S8 skewness = -1.90), we implemented a square root transformation. As expected, participants who perceived empathy to be more subjectively costly were less likely to choose empathy, $r = -.28$, $p < .001$. This result replicates the association from Study 5 reported in the main text

.

Table S11. NASA Task Load Index effort ratings, all studies.

Study	Emp-Obj M_{diff} [95% CI]	*p*	Emp-Obj Hedges' *g*	Choice *r*	*p* for Choice *r*	*N*
1. Labeled decks	0.99 [0.66, 1.32]	< .001	0.94	-.20	.148	56
2. Non-labeled decks	1.01 [0.73, 1.30]	< .001	0.90	-.32	.029	47
3. Empathy vs. emotion recog.	0.20 [0.09, 0.32]	< .001	0.17	-.26	< .001	196
4. Valence manip., between	0.60 [0.43, 0.77]	< .001	0.54	-.22	.002	193
5. Valence manip., between	0.70 [0.54, 0.86]	< .001	0.61	-.28	< .001	206
6. Valence manip., within	-0.02 [-0.31, .027]	.891	-0.02	.09	.525	50
7. Feel self/other, low arousal	0.46 [0.25, 0.66]	< .001	0.38	-.11	.291	91
8. Feel self/other, med. arousal	0.48 [0.26, 0.70]	< .001	0.42	-.21	.057	87
9. Efficacy manipulation	1.22 [0.85, 1.59]	< .001	1.11	-.42	< .001	90
10. Efficacy manipulation	1.77 [1.42, 2.12]	< .001	1.89	-.45	< .001	93
11. Varying empathic demand	0.14 [0.04, 0.24]	.007	0.13	-.04	.604	192
S1. Labeled decks	0.86 [0.65, 1.06]	< .001	0.75	-.04	.680	100
S2. Labeled decks	0.59 [0.35, 0.83]	< .001	0.49	-.05	.588	106
S3. Non-labeled decks	1.03 [0.74, 1.31]	< .001	0.87	-.01	.946	74
S4. Probabilistic decks	0.18 [-0.04, 0.39]	.108	0.17	-.06	.669	54
S5. One-word response	0.61 [0.44, 0.78]	< .001	0.52	-.17	.015	199
S7. Three-word response pilot	0.56 [-0.22, 1.33]	.147	0.51	-.73	.001	18
S8. Valence manip., between	0.37 [0.19, 0.55]	< .001	0.33	-.36	< .001	196
S9. Counterbalanced deck pos.	0.63 [0.47, 0.79]	< .001	0.52	-.29	< .001	200
S10. Valence manip., between	0.59 [0.42, 0.77]	< .001	0.55	-.39	< .001	201
S11. Vignette stimuli	0.10 [-0.06, 0.25]	.210	0.09	-.23	.001	188
Meta-analytic effect		< .001	**0.51**	**-.22**	< .001	2637

Note on Tables S11-S13. Mean difference computed as rating for empathy deck minus rating for objective deck. For deck comparisons, the low-efficacy conditions in Studies 9-10 are excluded because typical deck differences in these costs were expected to be reduced (N = 2539 for effort, N = 2538 for aversion and efficacy). For associations with empathy choice, these conditions are re-included because associations between choice and effort costs were expected to be similar across efficacy conditions (N = 2637 for effort, N = 2636 for aversion and efficacy). Values in the sample size column correspond to the analysis including high-efficacy conditions. The bottom row values are from random-effects meta-analyses.

Table S12. NASA Task Load Index aversion ratings, all studies.

Study	Emp-Obj M_{diff} [95% CI]	*p*	Emp-Obj Hedges' *g*	Choice *r*	*p* for Choice *r*	*N*
1. Labeled decks	0.71 [0.33, 1.10]	< .001	0.54	-.11	.403	56
2. Non-labeled decks	0.94 [0.56, 1.32]	< .001	0.72	-.16	.283	47
3. Empathy vs. emotion recog.	0.16 [0.03, 0.29]	.018	0.12	-.35	< .001	196
4. Valence manip., between	0.47 [0.30, 0.64]	< .001	0.33	-.17	.018	193
5. Valence manip., between	0.39 [0.22, 0.56]	< .001	0.28	-.24	< .001	206
6. Valence manip., within	0.12 [-0.22, 0.46]	.485	0.09	-.14	.316	50
7. Feel self/other, low arousal	0.44 [0.23, 0.65]	< .001	0.33	-.12	.244	91
8. Feel self/other, med. arousal	0.22 [-0.01, 0.44]	.058	0.19	-.18	.105	87
9. Efficacy manipulation	1.22 [0.78, 1.66]	< .001	1.19	-.55	< .001	90
10. Efficacy manipulation	1.33 [1.01, 1.65]	< .001	1.33	-.35	.001	93
11. Varying empathic demand	0.04 [-0.13, 0.21]	.675	0.03	-.08	.282	192
S1. Labeled decks	0.50 [0.23, 0.77]	< .001	0.36	-.06	.564	100
S2. Labeled decks	0.48 [0.24, 0.72]	< .001	0.33	-.13	.173	106
S3. Non-labeled decks	0.86 [0.52, 1.21]	< .001	0.63	-.31	.007	74
S4. Probabilistic decks	0.30 [0.07, 0.53]	.012	0.25	-.34	.011	54
S5. One-word response	0.44 [0.26, 0.62]	< .001	0.31	-.18	.010	198
S7. Three-word response pilot	0.33 [-0.76, 1.43]	.528	0.25	-.81	< .001	18
S8. Valence manip., between	0.01 [-0.18, 0.20]	.915	0.01	-.40	< .001	196
S9. Counterbalanced deck pos.	0.23 [0.05, 0.41]	.010	0.17	-.33	< .001	200
S10. Valence manip., between	0.40 [0.20, 0.60]	< .001	0.30	-.48	< .001	201
S11. Vignette stimuli	-0.13 [-0.30, 0.04]	.142	-0.10	-.23	.001	188
Meta-analytic effect		< .001	**0.30**	**-.27**	< .001	2636

Table S13. NASA Task Load Index efficacy ratings, all studies.

Study	Emp-Obj M_{diff} [95% CI]	*p*	Emp-Obj Hedges' *g*	Choice *r*	*p* for Choice *r*	*N*
1. Labeled decks	-0.38 [-0.74, -0.01]	.043	-0.34	.26	.056	56
2. Non-labeled decks	-0.74 [-1.08, -0.41]	< .001	-0.74	.36	.013	47
3. Empathy vs. emotion recog.	-0.14 [-0.28, -0.01]	.041	-0.14	.43	< .001	196
4. Valence manip., between	-0.55 [-0.71, -0.40]	< .001	-0.51	.50	< .001	193
5. Valence manip., between	-0.43 [-0.59, -0.28]	< .001	-0.41	.26	< .001	206
6. Valence manip., within	-0.43 [-0.87, 0.01]	.057	-0.36	.50	< .001	49
7. Feel self/other, low arousal	-0.52 [-0.73, -0.30]	< .001	-0.50	.29	.005	91
8. Feel self/other, med. arousal	-0.79 [-1.03, -0.56]	< .001	-0.77	.39	< .001	87
9. Efficacy manipulation	-1.35 [-1.74, -0.97]	< .001	-1.45	.59	< .001	90
10. Efficacy manipulation	-1.43 [-1.80, -1.06]	< .001	-1.57	.42	< .001	93
11. Varying empathic demand	-0.20 [-0.34, -0.06]	.007	-0.20	.21	.004	192
S1. Labeled decks	-0.19 [-0.41, 0.03]	.089	-0.17	.29	.003	100
S2. Labeled decks	-0.53 [-0.78, -0.27]	< .001	-0.45	.35	< .001	106
S3. Non-labeled decks	-0.46 [-0.76, -0.16]	.003	-0.41	.50	< .001	74
S4. Probabilistic decks	-0.09 [-0.25, 0.07]	.255	-0.11	.42	.001	54
S5. One-word response	-0.25 [-0.41, -0.08]	.004	-0.23	.44	< .001	199
S7. Three-word response pilot	-0.22 [-1.08, 0.64]	.594	-0.16	.75	< .001	18
S8. Valence manip., between	-0.12 [-0.29, 0.05]	.180	-0.11	.53	< .001	196
S9. Counterbalanced deck pos.	-0.17 [-0.34, -0.00]	.049	-0.16	.43	< .001	200
S10. Valence manip., between	-0.20 [-0.36, -0.04]	.017	-0.20	.39	< .001	201
S11. Vignette stimuli	0.27 [0.08, 0.45]	.005	0.26	.20	.005	188
Meta-analytic effect		< .001	**-0.35**	**.40**	< .001	2636

Table S14. Multiple regression analyses for NASA Task Load Index cost ratings.

	Effort		Aversion		Efficacy	
Study	β	*p*	β	*p*	β	*p*
1. Labeled decks	-.22	.202	.12	.498	.26	.077
2. Non-labeled decks	-.19	.238	-.09	.541	.30	.044
3. Empathy vs. emotion recog.	-.05	.551	-.16	.053	.34	< .001
4. Valence manip., between	-.13	.075	.03	.675	.48	< .001
5. Valence manip., between	-.20	.007	-.09	.239	.20	.005
6. Valence manip., within	.04	.795	-.02	.916	.49	.001
7. Feel self/other, low arousal	.05	.728	-.10	.456	.29	.009
8. Feel self/other, med. arousal	-.02	.884	-.06	.584	.37	.002
9. Efficacy manipulation	-.12	.318	-.16	.288	.41	.002
10. Efficacy manipulation	-.33	.028	.06	.679	.23	.125
11. Varying empathic demand	.00	.998	-.06	.428	.20	.005
S1. Labeled decks	-.07	.484	-.02	.845	.30	.003
S2. Labeled decks	.04	.745	-.05	.646	.35	.001
S3. Non-labeled decks	.18	.169	-.32	.015	.43	< .001
S4. Probabilistic decks	-.01	.917	-.24	.089	.36	.009
S5. One-word response	-.01	.881	-.12	.169	.42	< .001
S7. Three-word response pilot	.19	.583	-.68	.041	.41	.095
S8. Valence manip., between	-.16	.034	-.09	.266	.43	< .001
S9. Counterbalanced deck pos.	-.04	.564	-.21	.006	.36	< .001
S10. Valence manip., between	-.16	.029	-.30	< .001	.21	.002
S11. Vignette stimuli	-.15	.056	-.13	.123	.14	.062

Note. Mean difference computed as rating for empathy deck minus rating for objective deck.

Table S15. Moderation of Time Course Effects in Empathy Selection Task by NASA Effort Ratings, Controlling for NASA Efficacy Ratings, Studies 1-11.

	B	SE	*t*	*p*	Exp(B)	95% CI Exp (B)
Study 1	-.02	.01	-2.06	.040	0.98	0.96, 1.00
Study 2	-.04	.02	-2.13	.033	0.96	0.93, 1.00
Study 3	-.03	.01	-3.45	.001	0.97	0.96, 0.99
Study 4	-.02	.01	-3.26	.001	0.98	0.97, 0.99
Study 5	-.02	.01	-3.24	.001	0.99	0.98, 0.99
Study 6	-.01	.02	-0.51	.607	0.99	0.96, 1.02
Study 7	-.00	.01	-0.31	.760	1.00	0.99, 1.01
Study 8	-.01	.01	-2.21	.027	0.99	0.97, 1.00
Study 9	-.01	.01	-1.21	.228	0.99	0.97, 1.01
Study 10	.01	.01	1.03	.305	1.01	0.99, 1.03
Study 11	-.02	.01	-1.59	.112	0.98	0.97, 1.00

Note. Results are from generalized linear mixed models with a random intercept and random slope for time, and an autoregressive covariance parameter. Regression coefficients are for the interaction between mean-centered NASA Effort difference scores (empathy deck minus objective deck) and the centered question order variable (with values ranging from 0-39 in

Studies 1-8 and 0-23 in Studies 9-10). The high-efficacy conditions of Studies 9 and 10 are excluded because empathy avoidance was not predicted to occur in these conditions. These analyses control for efficacy scores.

Table S16. Relationships between NASA Task Load Index measures, all studies.

	Effort-Aversion		Effort-Efficacy		Aversion-Efficacy	
	r	*p*	*r*	*p*	*r*	*p*
1. Labeled decks	.65	< .001	-.21	.126	-.37	.005
2. Non-labeled decks	.39	.007	-.30	.041	.03	.861
3. Empathy vs. emotion recog.	.58	< .001	-.33	< .001	-.47	< .001
4. Valence manip., between	.47	< .001	-.22	.002	-.29	< .001
5. Valence manip., between	.45	< .001	-.18	.010	-.31	< .001
6. Valence manip., within	.55	< .001	.14	.357	-.28	.051
7. Feel self/other, low arousal	.60	< .001	-.35	.001	-.19	.068
8. Feel self/other, med. arousal	.40	< .001	-.44	< .001	-.29	.006
9. Efficacy manipulation	.69	< .001	-.47	< .001	-.74	< .001
10. Efficacy manipulation	.74	< .001	-.71	< .001	-.73	< .001
11. Varying empathic demand	.35	< .001	-.08	.267	-.09	.239
S1. Labeled decks	.34	.001	.13	.202	-.04	.668
S2. Labeled decks	.58	< .001	-.17	.080	-.29	.002
S3. Non-labeled decks	.60	< .001	.03	.834	-.21	.072
S4. Probabilistic decks	.30	.025	.08	.585	-.29	.036
S5. One-word response	.65	< .001	-.21	.003	-.14	.049
S7. Three-word response pilot	.88	< .001	-.78	< .001	-.72	.001
S8. Valence manip., between	.59	< .001	-.35	< .001	-.51	< .001
S9. Counterbalanced deck pos.	.52	< .001	-.39	< .001	-.28	< .001
S10. Valence manip., between	.54	< .001	-.30	< .001	-.43	< .001
S11. Vignette stimuli	.44	< .001	-.17	.017	-.30	< .001

Note. Cost measures reflect difference scores (empathy deck minus objective deck).