Supplementary Materials

Study S1: Norming for Moral Distinctiveness

Norming Study S1 sought to ensure that participants perceived the vignettes used in the current studies as distinctly moral, in opposition to the conventional decision-making dilemmas used in previous research (e.g., Johnson & Rips, 2015).

Methods. We recruited 49 participants (M_{age} = 36, 54% female); 22 were excluded due to incorrect answers to check questions.

Participants read thirteen vignettes in a random order — the eight vignettes used in the current studies and the five vignettes used in a previous study of lay decision theory (Johnson & Rips, 2015), such as the item concerning Jill's choice of shampoo described in the main text. After the description of each dilemma, participants were told that the agent deliberately chose the Worst option (e.g., the doctor deliberately chooses the option most likely to lead to permanent hearing loss, or Jill deliberately chooses the option least likely to make her hair smell like apples). Participants then rated the following (in this order) on scales from 1 to 9:

Seriousness. "How seriously wrong is [the doctor's] action?"

Universality. "Is it wrong in other places and times?"

Authority-independence. "Would it still be wrong if someone in authority said it was OK?"

Objectivity. "Imagine that someone else disagrees with you about whether the action was wrong. Must one of you be wrong?"

Harm. "Does the action lead to serious harm?"

Items were presented in a random order.

Results and discussion. On every measure, all of the morally laden vignettes were rated above the scale midpoint and all of the morally neutral vignettes were rated below the scale midpoint. These differences were significant for all measures: wrongness [M = 8.48, SD = 0.73 vs. M = 3.38, SD = 2.03; $t(26) = 12.23$, $p < .001$], universality [M = 8.27, SD = 0.93 vs. M = 3.30, SD = 2.05; $t(26) = 12.00$, $p < .001$], authority [M = 7.86, SD = 1.61 vs. M = 3.21, SD = 2.11; $t(26) = 9.75$, $p < .001$], objectivity [M = 7.84, SD = 1.28 vs. M = 3.53, SD = 2.47; $t(26) = 9.85$, $p < .001$], and harm [M = 8.30, SD = 0.60 vs. M = 1.87, SD = 1.40; $t(26) = 21.87$, $p < .001$]. Thus, the vignettes used in the current study were distinctly moral.

Study S2: Exploring Differences Among Vignettes

The purpose of Norming Study S2 was to test whether the seriousness of the harm invoked in each vignette — measured in terms of severity, directness, and number harmed — can explain differences in the size of the optimality bias across vignettes.

Methods. We recruited 38 participants (M_{age} = 32, 45% female); 4 were excluded due to incorrect answers to check questions.

Participants read the eight vignettes used in the current studies. After a description of each dilemma, participants were told that the agent deliberately chose the Worst option, as in Norming Study S1. Participants then rated the following (in this order) on scales from 1 to 9:

Severity. "How serious were the consequences of the [doctor's] actions?"

Directness. "Did the [doctor's] actions lead to direct, bodily harm?"

Number. "How many people were harmed because of the [doctor's] actions?"

Items were presented in a random order.

	Best–Middle	Middle–Worst	Severity	Directness	Number
Doctor	1.20***	0.30	8.24	7.94	1.62
Farmer	0.75*	0.19	8.44	8.24	3.97
Building Contractor	1.08***	0.21	8.68	8.68	2.79
Computer Programmer	1.42***	0.12	8.09	2.09	5.03
Jet Pilot	1.07***	0.10	8.68	8.68	2.71
Paramedic	0.92***	0.48°	8.76	8.38	1.79
CEO	1.49***	0.17	8.65	8.53	3.00
Investment Banker	0.96**	0.02	8.29	2.21	1.79

° $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table S1. Item characteristics for each vignette.

Note. The first two columns give difference scores for each vignette, pooling data across studies, while the last three columns give the measurements of moral seriousness from Study S2. Significance levels are indicated for the Best–Middle and Middle–Worst comparisons.

Results and discussion. The vignette means for each measure are shown in Table S1, along with the mean differences for each comparison (Best–Middle, and Middle–Worst). The latter data were obtained by selecting all participants across Studies 1–5 in the main text, as well as supplementary Studies S3A, S3B, S4A, S4B, S5A, S5B, and S6 (the same studies used in the meta-analysis presented in the main text). The 7-point scales used in Study 1 were rescaled to a 9-point scale for consistency. The difference between Best and Middle was significant for all 8 vignettes. The difference between Middle and Worst reached marginal significance for only one out of the 7 vignettes, but was always numerically positive. This pattern is broadly consistent with the meta-analysis presented in the main text, which revealed robust differences between the Best and Middle conditions but only very small differences between the Middle and Worst conditions.

The Best–Middle difference — the measure of the optimality bias — was not significantly associated with severity [$r(6) = –.24$, $p = .56$], directness [$r(6) = –.19$, $p = .66$], or number [$r(6) = .28$, $p = .50$]. A multiple regression, adjusting for the effects of these variables simultaneously, reaches similar conclusions. Of course, strong conclusions about item differences cannot be drawn from a study of 8 vignettes.

The Middle–Worst difference also was not significantly associated with severity [$r(6) = .40$, $p = .33$], directness [$r(6) = .53$, $p = .17$], or number [$r(6) = –.35$, $p = .40$]. Although these correlations are not highly reliable, they are fairly large in magnitude. Perhaps when the harm is sufficiently serious, people are more inclined to attend not only to optimality, but also to more fine-grained probability information. This conclusion cannot be safely drawn from the current studies — particularly since the difference between Middle and Worst did not reach significance for most of the vignettes, taken individually — but is suggestive for future research.

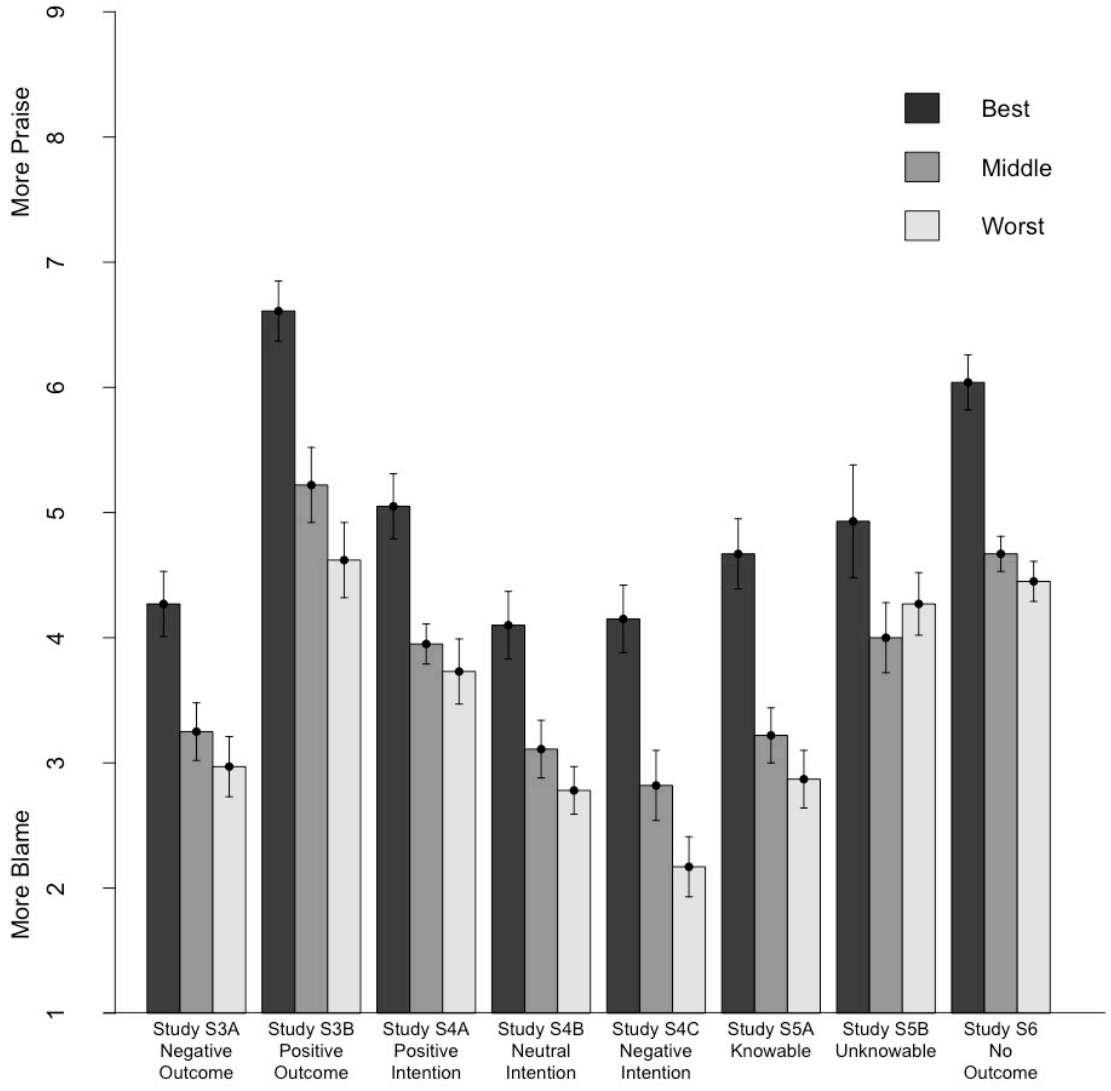

Figure S1. Results of Studies S3–S6.

Note. Bars represent 1 SE. Scales reverse-coded.

Study S3: Positive and Negative Outcomes

Although Study 4 demonstrated the optimality bias for positive outcomes, it did so in an unusual context where the outcomes are inevitable. Study S3 directly contrasted positive and negative outcomes in the more common situation where the outcomes are uncertain *ex ante*. In addition, this study allowed us to compare the magnitude of the *outcome bias* — the tendency to make harsher moral evaluations in light of negative outcomes (Baron & Hershey, 1988) — to that of the optimality bias.

Methods. We recruited 267 participants (M_{age} = 32, 39% female); 67 were excluded due to incorrect answers to check questions.

The method was the same as those in Study 1 in the main text (see Appendices S1–3), with three changes. First, the outcome was negative for some participants (Study S3A) and positive for others (Study S3B). Second, the information about the agent having done due diligence (used in Studies 2–7) was included. Finally, the dependent measure was moral blame (as in Studies 2–7).

Results and discussion. When judging the blameworthiness of the agents' actions for the negative outcomes in Study S3A, participants blamed agents choosing Best less than those choosing Middle [$t(64)$ = 2.86, p = .006, d = 0.71, 95% CI[0.31,1.73], BF_{10} = 6.5, d_S = 0.74; Figure S1], but blamed agents choosing Middle no less than those choosing Worst [$t(71)$ = 0.82, p = .42, d = 0.19, 95% CI[–0.40,0.95], BF_{01} = 4.1, d_S = 0.67]. Likewise, for the positive outcomes of Study S3B, moral judgments again tracked optimality, with more praise assigned to agents choosing Best than Middle [$t(61)$ = 3.57, p = .001, d = 0.90, 95% CI[0.61,2.18], BF_{10} = 40.2, d_S = 0.72], but no more praise assigned to agents choosing Middle than Worst [$t(64)$ = 1.43, p = .16, d = 0.35, 95% CI[–0.24,1.44], BF_{01} = 2.1, d_S = 0.71]. The interaction between experiment and condition was not significant [$F(2,194)$ = 0.85, p = .43], failing to support a moderating effect of outcome. Hence, the same efficiency-based mechanisms appear to apply to moral judgments made in light of *both* negative and positive outcomes.

Although the patterns were similar for positive and negative outcomes, participants generally produced harsher evaluations of an agent's behavior in light of negative outcomes (Study S3A) than positive outcomes (Study S3B), even though the agent could control the outcome only indirectly through her choice [$t(198) = 8.47$, $p < .001$, $d = 1.20$, 95% CI[1.54,2.47], $BF_{10} > 1000$, $d_S = 0.39$], demonstrating a robust *outcome bias* (Baron & Hershey, 1988). Notably, the size of the optimality bias ($d = 0.71$ for negative outcomes and $d = 0.90$ for positive outcomes) approached the size of the outcome bias ($d = 1.20$).

Like Study 1, the omission of the unknowability stipulation from these stimuli (as well as those used in Study S4) raises important questions of internal validity: Could participants have believed that the agents did their research poorly, and hence blamed them for their ignorance? In the absence of strong stipulations about unknowability, this remains a possibility. Nonetheless, finding the same pattern as in the main experiments for both positive and negative outcomes suggests that the effect generalizes to cases where such stipulations are not made, which may better reflect the real world.

Study S4: Positive, Neutral, and Negative Intentions

One possible concern is that participants view the agents' choices as reflecting their *goals* — optimal decisions may have signaled a positive intention, whereas suboptimal decisions may not have. In that case, the increased blame for the suboptimal agents would reflect blame for the agent's mental states, not for their choice (although it is difficult to see, normatively, how the agents' positive or negative intentions could have driven which of the three options they chose, given their ignorance of the probabilities associated with these options). To address this possibility, Study S4A specified that the agent had a positive intention, and Study S4B specified that the agent did not have a positive intention (i.e., a "neutral" intention).

In addition, we examined a potential consequence of our efficiency account. From the moral *patient's* point of view, the agent's optimal choice is always Best. However, from the *agent's* perspective, the optimal choice depends on their intention. For a positively intentioned agent, the optimal choice is Best. But if the agent has negative intentions, the Worst option is actually optimal. Thus, perhaps when the agent has a

negative intention, participants would distinguish between Middle and Worst — options that are always suboptimal from the patient's perspective, but which differ in optimality from the agent's perspective. Study S4C tested this possibility by specifying that the agent had a negative intention.

Methods. We recruited 503 participants (M_{age} = 34, 44% female); 202 were excluded because they incorrectly answered one or more check questions.

The method was the same as Study S3A, except that the agent's intent was specified to be either positive (e.g., "The doctor intends to choose the best treatment option for her patient"; Study S4A), neutral (e.g., "The doctor does not intend to choose the best treatment option for her patient"; Study S4B), or negative (e.g., "The doctor intends to choose the worst treatment option for her patient"; Study S4C). The outcome of the agent's action was always negative.

Results and discussion. Given positive intentions in Study S4A**,** judgments of blame for the positive outcome again tracked optimality, since agents were evaluated more positively if they chose Best than Middle [$t(76) = 3.54$, $p = .001$, $d = 0.80$, 95% CI[0.48,1.72], $BF_{10} = 39.5$, $d_S = 0.65$; Figure S1], but no differently if they chose Middle rather than Worst [$t(68) = 0.76$, $p = .45$, $d = 0.18$, 95% CI[–0.36,0.79], $BF_{01} = 4.2$, $d_S = 0.63$]. Similarly, given neutral intentions in Study S4B, judgments of blame tracked optimality, since agents were evaluated more positively if they chose Best than Middle [$t(65) = 2.64$, $p = .010$, $d = 0.65$, 95% CI[0.24,1.74], $BF_{10} = 4.0$, $d_S = 0.74$] but no differently if they chose Middle rather than Worst [$t(71) = 1.09$, $p = .28$, $d = 0.25$, 95% CI[–0.28,0.94], $BF_{01} = 3.3$, $d_S = 0.66$]. Thus, an optimal choice does not function merely as a signal of the agent's unrevealed intention, since the effect occurs even if the intention is specified. Comparing the means for Study S4A and S4B reveals a sizeable effect of positive versus neutral intention [$t(209) = 4.83$, $p < .001$, $d = 0.67$, 95% CI[0.59,1.40], $BF_{10} > 1000$, $d_S = 0.38$]. However, the effect size of the optimality bias was, if anything, even larger ($d = 0.80$ and $d = 0.65$) than the effect of intention ($d = 0.67$).

Unlike for positive and neutral intentions, for negative intentions there is a mismatch in optimality — Worst is the optimal choice from the agent's perspective, whereas Best is the optimal choice from the victim's perspective. Therefore, we predicted (*a priori*) that this clash in optimalities might cause some sign of a stepwise assignments of blame, unlike in the other conditions and the previous studies. Providing some evidence for this prediction, participants in Study S4C judged agents with negative intentions as more blameworthy when choosing Middle than Best [$t(53) = 2.84$, $p = .006$, $d = 0.76$, 95% CI[0.39,2.27], $BF_{10} = 6.2$, $d_S = 0.78$], but also marginally more blameworthy when choosing Worst than Middle [$t(61) = 1.78$, $p = .081$, $d = 0.45$, 95% CI[–0.08,1.38], $BF_{01} = 1.3$, $d_S = 0.76$]. However, the interaction between condition and experiment was not significant [$F(4,292) = 0.54$, $p = .70$], failing to support a moderating effect of intention.

One possibility is that the manipulation of intention was unsuccessful because participants drew exculpatory inferences about why the agents chose the worst option (e.g., maybe the patient desired that the doctor act the way she did), although this seems unlikely for many of the vignettes (e.g., it is not very plausible that the patient would ask for permanent hearing loss). Whatever the reason for these mixed results, they do not allow strong conclusions about intention-based differences in the optimality bias when the agent and patient have *different* optimal choices. Future research might further address the potential role of motivations on moral judgments under these particular circumstances.

Study S5: Knowable and Unknowable Probabilities

Studies 2–6 addressed the issue of perceived negligence (Nobes, Panagiotaki, & Pawson, 2009) in two ways — specifying in all cases that the agent did due diligence to assess the probabilities of the choices and that these probabilities were unknowable, and directly measuring attributions of negligence in Study 5. In Study S5, we further investigated the role of perceived negligence by comparing cases where the probabilities were unknown but *knowable* to cases where the probabilities were both unknown and *unknowable*.

Methods. We recruited 335 participants ($M_{age} = 32$, 37% female); 188 were excluded due to incorrect

answers to check questions.

The vignettes were the same as those in Study S3A, except that participants were told either that evidence existed indicating that the options had different likelihoods of a positive outcome (*knowable* condition, e.g., "In fact, there is some existing evidence that says this belief is incorrect. As it happens, the doctor's belief is wrong."; Study S5A) or that no such evidence existed (*unknowable* condition, e.g., "In fact, all of the existing evidence says that this belief is correct. But as it happens, for reasons completely outside of her control, the doctor's belief is wrong."; Study S5B).

Results and discussion. In Study S5A, when the probabilities were knowable, participants judged the agent less harshly when she chose Best rather than Middle [$t(70) = 4.46$, $p < .001$, $d = 1.05$, 95% CI[0.80,2.09], $BF_{10} = 656.1$, $d_S = 0.67$; Figure S1], but equally harshly whether she chose Middle or Worst [$t(65) = 1.10$, $p = .27$, $d = 0.27$, 95% CI[–0.28,0.99], $BF_{01} = 3.1$, $d_S = 0.67$]. Likewise, in Study S5B, when the probabilities were unknowable, participants again judged the agent less harshly when she chose Best rather than Middle [$t(27) = 2.20$, $p = .037$, $d = 0.82$, 95% CI[0.06,1.80], $BF_{10} = 1.8$, $d_S = 1.10$], but equally harshly whether she chose Middle or Worst [$t(28) = –0.72$, $p = .48$, $d = –0.26$, 95% CI[–1.03,0.49], $BF_{01} = 3.1$, $d_S = 1.06$]. The interaction between condition and experiment was not significant [$F(2,141) = 2.02$, $p = .14$], failing to support a moderating effect of knowability. Thus, participants in the other studies do not seem to have been blaming suboptimal agents for their negligence in failing to know the probabilities, but rather for choosing suboptimally.

The effect of optimality was similar regardless of whether the outcome probabilities were knowable or unknowable. However, participants made harsher judgments overall when the probabilities were knowable (Study S5A) rather than unknowable (Study S5B) [$t(145) = 2.97$, $p = .003$, $d = 0.53$, 95% CI[0.26,1.27], $BF_{10} = 8.3$, $d_S = 0.39$]. This suggests that people expect moral agents to make exhaustive efforts to inform themselves about their decision, and will hold them accountable for not doing so. However, the size of this

negligence effect (d = 0.53) was only about half that of the optimality effect (d = 1.05 for Study S5A and d = 0.82 for Study S5B).

Study S6: Unknown Outcomes

Study S6 tested two further questions about efficiency-based moral judgment. First, would this effect occur even when the outcome is unknown? Given that moral judgments are often tied to concrete outcomes (e.g., Baron & Hershey, 1988), people may only judge the agent's suboptimal choice relative to a particular outcome. This can lead to a biased search for justifications in light of the outcome (see Alicke, 2000). When the outcome is negative, people may search for reasons to justify negative judgments (and suboptimality would be one such reason) and when the outcome is positive, people may search for reasons to justify positive judgments (and optimality would again be one such reason). But when there is no outcome, this sort of outcome-based reasoning would be avoided altogether, which might erase the optimality bias (see Tversky & Shafir, 1992 for a similar phenomenon). Study S6 tested this possibility by omitting the outcome.

Second, to what extent do participants have conscious awareness of using efficiency-based thinking? Study 7 in the main text found that people mispredict their blame judgments for ignorant agents, believing that they would either differentiate among all three choices or give similar judgments to all three choices. This suggests that people lack awareness of using the Efficiency Principle. To further test this issue, participants in Study S6 were asked to choose a justification for their judgments.

Methods. We recruited 336 participants (M_{age} = 33, 45% female); 90 were excluded due to incorrect answers to check questions.

The method was the same as Study S5B, with two changes. First, the outcome was omitted from the vignettes. Second, on the page after the main measures, participants were asked to justify their responses by selecting one of three options — a mentalistic justification ("The doctor believed that she gave her patient

the best treatment available, and couldn't have known otherwise, therefore she is not responsible for the harm which this choice may cause"), an efficiency-based justification (e.g., "The fact that the doctor didn't know which treatment was best does not remove her responsibility for the harm which this choice may cause"), or 'other' (writing in their own justification). The order of the mentalistic and efficiency-based justifications was randomized.

Results and discussion. Although Study S6 omitted outcomes, judgments were again more positive for agents choosing Best rather than Middle [$t(169) = 6.49$, $p < .001$, $d = 0.99$, 95% CI[0.95,1.79], $BF_{10} > 1000$, $d_S = 0.44$; Figure S1], but did not differ between Middle and Worst [$t(163) = 1.01$, $p = .32$, $d = 0.16$, 95% CI[–0.20,0.63], $BF_{01} = 5.0$, $d_S = 0.42$]. Thus, efficiency-based judgments are not tied to concrete outcomes, but occur even when the outcome is not specified. Indeed, the size of the effect was no smaller when outcomes were omitted ($d = 0.99$) as compared to previous studies (varying from $d = 0.30$ to $d = 1.11$), suggesting that efficiency-based thinking is tied entirely to the *optimality of the agent's decision* rather than to the outcome.

Justification	Proportion	Means		
		Best	Middle	Worst
Mentalistic	65.0%	6.27	5.28	5.13
Efficiency-Based	27.6%	5.06	3.43	3.29
Other	7.3%	6.30	4.50	3.50

Table S2. Response patterns in Study S6.

Note. Means are broken down according to the justifications chosen by participants. Means for the 'other' justifications are highly imprecise given the small cell sizes (18 participants total across three between-subjects conditions).

Most participants appear to be unaware of using an efficiency-based strategy, because 65.0% of participants selected the mentalistic justification (which argued that the agent should be exculpated), compared to 27.6% who selected the efficiency-based justification (which held the agent blameworthy) and 7.3% who selected

'other' (see Table S2). Further, responses were similar regardless of the explicit justification provided: Both the mentalistic and efficiency-based participants distinguished between Best and Middle [$t(111) = 4.64$, $p < .001$, $d = 0.87$, 95% CI[0.57,1.42], $BF_{10} > 1000$, $d_S = 0.54$ for mentalistic; $t(42) = 3.88$, $p < .001$, $d = 1.22$, 95% CI[0.78,2.48], $BF_{10} = 77.2$, $d_S = 1.02$ for efficiency-based], while neither group distinguished between Middle and Worst [$t(103) = 0.74$, $p = .46$, $d = 0.15$, 95% CI[–0.25,0.54], $BF_{01} = 5.1$, $d_S = 0.52$ and $t(50) = 0.45$, $p = .66$, $d = 0.13$, 95% CI[–0.47,0.75], $BF_{01} = 4.4$, $d_S = 0.76$, respectively]. Thus, even participants who self-reported an exculpatory mentalistic justification used efficiency-based reasoning.

These results suggest, together with Study 7, that participants lacked introspective access to the effects of the Efficiency Principle on their behavior. Further, people appear to construct post hoc justifications for their efficiency-based inferences, since nearly two-thirds of participants chose the mentalistic justification. Although it is possible that some participants thought that both justifications had merit, the key point is that regardless of which justification was deemed more important, participants had the same degree of optimality bias. This makes the bias all the more troubling in everyday moral judgment, since people may have difficulty identifying when they or others are falling prey to it.

Appendix S1: Vignette Wordings

Note. Exact text corresponds to Study 1 in the main text. Cross-reference with Appendices S2 and S3 to construct the vignettes used in the other studies.

Doctor A doctor working in a hospital has a patient who is having hearing problems. This patient has three, and only three, treatment options. The doctor believes that all treatment options have a 70% chance of giving the patient a full, successful recovery. But in fact the doctor's belief is wrong. Actually:

1) If she gives the patient treatment LPN, there is a 70% chance the patient will have a full recovery.
2) If she gives the patient treatment PTY, there is a 50% chance the patient will have a full recovery.
3) If she gives the patient treatment NRW, there is a 30% chance the patient will have a full recovery.

The doctor chooses treatment LPN, and the patient does not recover at all. The patient now has permanent hearing loss.

Farmer A local farmer needs to add insecticide to all her crops. The farmer has three, and only three, insecticide options. The farmer believes that all options have a 70% chance of solving her insect problem while also leaving the local river completely uncontaminated. But in fact the farmer's belief is wrong. Actually:

1) If she uses brand LPN, there is a 70% chance the river will remain uncontaminated.

2) If she uses brand PTY, there is a 50% chance the river will remain uncontaminated.
3) If she uses brand NRW, there is a 30% chance the river will remain uncontaminated.

The farmer chooses option [LPN/PTY/NRW], and the river becomes contaminated. A few locals drink the water and get extremely sick.

Contractor A building contractor is deciding what cement to use to build the foundation of a new house. The contractor has three, and only three, cement options. The contractor believes that all options have a 70% chance of withstanding an earthquake, if one occurs. But in fact the contractor is wrong. Actually:

1) If he uses brand LPN, there is a 70% chance the house will withstand an earthquake.
2) If he uses brand PTY, there is a 50% chance the house will withstand an earthquake.
3) If he uses brand NRW, there is a 30% chance the house will withstand an earthquake.

The contractor chooses brand [LPN/PTY/NRW], and one day a brief earthquake occurs and the house very quickly collapses. Two people who were in the house are killed.

Programmer A computer programmer is deciding what anti-virus software to use in order to protect top-secret government information. The programmer has three, and only three, software options. The programmer believes that all software options have a 70% chance of saving the government's information in the unlikely event that a virus penetrates the system. But in fact the programmer is wrong. Actually:

1) If she uses software LPN, there is a 70% chance that the software will be able to save the information.
2) If she uses software PTY, there is a 50% chance that the software will be able to save the information.
3) If she uses software NRW, there is a 30% chance that the software will be able to save the information.

The programmer chooses software [LPN/PTY/NRW], and 5 months later a virus penetrates the system and all the information is immediately lost.

Jet Pilot A jet pilot is deciding what type of missile to use for a combat mission that will take place at a very high altitude. The pilot has three, and only three, missile options. The pilot believes that all missile options have a 70% chance of successfully hitting their target at this altitude. But in fact the pilot is wrong. Actually:

1) If he uses missile LPN, there is a 70% chance the missile will successfully hit its target.
2) If he uses missile PTY, there is a 50% chance the missile will successfully hit its target.
3) If he uses missile NRW, there is a 30% chance the missile will successfully hit its target.

The pilot chooses missile [LPN/PTY/NRW], and during combat the missile completely misses its target and instead hits a jet from the pilot's own air force. The pilots of that jet are immediately killed.

Paramedic A paramedic needs to drive to a hospital in order to deliver a kidney to a patient who is having an emergency kidney transplant. The paramedic has three, and only three route options to get to the hospital. The paramedic believes that all the options will give him a 70% chance of arriving in time to save the patient's life. But in fact the paramedic is wrong. Actually:

1) If he takes route LPN, there is a 70% chance he will successfully deliver the kidney in time.
2) If he takes route PTY, there is a 50% chance he will successfully deliver the kidney in time.
3) If he takes route NRW, there is a 30% chance he will successfully deliver the kidney in time.

The paramedic chooses route [LPN/PTY/NRW], and arrives far too late. The patient dies.

CEO The CEO of a company is deciding what type of material to use in order to make a new biohazard suite. The CEO has three, and only three options. The CEO believes that all three options have a 70% chance of protecting suit users from high levels of radiation. But in fact the CEO is wrong. Actually:

1) If he chooses material LPN, there is a 70% chance that suit users will be protected from high levels of radiation.
2) If he chooses material PTY, there is a 50% chance that suit users will be protected from high levels of radiation.
3) If he chooses material NRW, there is a 30% chance that suit users will be protected from high levels of radiation.

The CEO chooses material [LPN/PTY/NRW], and 2 months later 2 suit users die of radiation poisoning.

Broker An investment broker is deciding which investment option to choose for her client's life savings. The investment broker has three, and only three, options. The investment broker believes that all three options have a 70% chance of surviving an economic downturn. But in fact the investment broker is wrong. Actually:

1) If she chooses option LPN, there is a 70% chance her client's life savings will survive an economic downturn.
2) If she chooses option PTY, there is a 50% chance her client's life savings will survive an economic downturn.
3) If she chooses option NRW, there is a 30% chance her client's life savings will survive an economic downturn.

The investment broker chooses option [LPN/PTY/NRW], and 3 months later there is a brief economic downturn and her client immediately loses her entire life's savings.

Appendix S2: Condition Wordings

Note. Cross-reference with Appendices S1 and S3 to construct the vignettes used in each study.

A doctor working in a hospital has a patient who is having hearing problems. This patient has three, and only three, treatment options.

Positive intention	The doctor intends to choose the best treatment option for her patient.
Neutral intention	The doctor does not intend to choose the best treatment option for her patient.
Negative Intention	The doctor intends to choose the worst treatment option for her patient.

Ignorant	The doctor believes that all treatment options have a 70% chance of giving the patient a full, successful recovery.
Due diligence	Based on many articles that the doctor has carefully read in respected medical journals, she truly believes that all three options have a 70% chance of giving the patient a full, successful recovery.
Knowledgeable	The doctor knows that:

Knowability unspecified	But in fact the doctor's belief is wrong. Actually:
Knowable	In fact, there is some existing evidence that says this belief is incorrect. As it happens, the doctor's belief is wrong. Actually:
Unknowable	In fact, all of the existing evidence says that this belief is correct. But as it happens, for reasons completely outside of her control, the doctor's belief is wrong. Actually:

1) If she gives the patient treatment LPN, there is a 70% chance the patient will have a full recovery.
2) If she gives the patient treatment PTY, there is a 50% chance the patient will have a full recovery.
3) If she gives the patient treatment NRW, there is a 30% chance the patient will have a full recovery.

Negative outcome	The doctor chooses treatment [LPN/PTY/NRW], and the patient does not recover at all. The patient now has permanent hearing loss.

Positive outcome	The doctor chooses treatment [LPN/PTY/NRW], and the patient recovers. The patient has no permanent hearing loss.
Unspecified outcome	The doctor chooses treatment [LPN/PTY/NRW].
Hypothetical negative outcome	Imagine that the doctor chose treatment [LPN/PTY/NRW], and that the patient did not recover at all, suffering permanent hearing loss.
Inevitable	We now also know that the patient had a gene that would have allowed any treatment to cure the disease.
Wrongness	How wrong was the doctor's behavior?
Punishment	How much should the doctor be punished?
Praise/blame	What does the doctor deserve to receive for her behavior?
Explanation	To what extent do you feel that an explanation is necessary for the doctor's choice?
Negligence	While answering the question about blame, did you think that if the doctor had thought more carefully or done more research, then she would have been able to know which options were better and which were worse?
Egocentrism	While answering the question about blame, did you think that the doctor had some sense of which options were better and which were worse?
Hypothetical praise/blame	What would the doctor deserve to receive for her behavior?

Appendix S3: Structure of Cases

Note. Cross-reference with Appendices S1 and S2 to construct the vignettes used in each study.

Study	Intent	Knowledge	Knowable	Outcome	Dependent Measures
1A	—	Ignorant	—	Negative	Wrongness
1B	—	Ignorant	—	Negative	Punishment
2	—	Due Diligence	Unknowable	Negative	Praise/Blame
3	—	Due Diligence	Unknowable	Negative	Explanation Praise/Blame
4	—	Due Diligence	Unknowable	Positive / Inevitable	Praise/Blame
5	—	Due Diligence	Unknowable	Negative	Praise/Blame Negligence Egocentrism
6A	—	Due Diligence	Unknowable	Negative	Wrongness

					Praise/Blame
6B	—	Knowledgeable	—	Negative	Wrongness Praise/Blame
7	—	Due Diligence	Unknowable	Hypothetical Negative	Hypothetical Praise/Blame
S3A	—	Due Diligence	—	Negative	Praise/Blame
S3B	—	Due Diligence	—	Positive	Praise/Blame
S4A	Positive	Due Diligence	—	Negative	Praise/Blame
S4B	Neutral	Due Diligence	—	Negative	Praise/Blame
S4C	Negative	Due Diligence	—	Negative	Praise/Blame
S5A	—	Due Diligence	Knowable	Negative	Praise/Blame
S5B	—	Due Diligence	Unknowable	Negative	Praise/Blame
S6	—	Due Diligence	Unknowable	Unspecified	Praise/Blame

Appendix S4: Sample Size Planning

Sample sizes for all experiments were planned *a priori*, using effect size estimates based on related studies (Johnson & Rips, 2014, *under review*). For most of the between-subjects studies (Studies 1, 3, 4 and S3–S5), we aimed for a sample size of 168 per study, in order to achieve 90% power. A somewhat smaller sample size was used in Study 3, due to a participant recruitment error.

We doubled the planned sample size (target $N = 336$, with small fluctuations across studies) for Studies 2, 5, and S6, for the following reasons. In the case of Study 2, a pilot study found a similar pattern of means but was underpowered due to a smaller effect size (compared to the other studies). In the case of Studies 5 and S6, we planned to study the relationship between participants' moral judgments and their responses to other questions (egocentrism and negligence attributions in Study 5 and justifications in Study S6), requiring a larger sample size.

We planned a smaller sample size (target $N = 96$, with small fluctuations across studies) in Studies 6 and 7, compared to the between-subjects studies. The studies manipulated the key independent variable within-subjects, so fewer participants were needed to achieve similar levels of power.

Sample sizes varied slightly from the target in some studies due to recruitment procedures. Sensitivity power analyses are reported for all *t*-tests, with the d_s statistics referring to the minimum effect size that would be detectable with 80% power given the actual sample size (after exclusions).

Appendix S5: Exclusion Criteria

Check Questions. Check questions were used in all studies in the main article and Supplementary Materials. For Studies 2–5, participants answered the following questions for their assigned vignette:

Belief. "TRUE or FALSE: The doctor believed that all [three treatments] had a 70% chance of leading to [recovery]."

Choice. "Given the treatment that the doctor chose, what was the actual chance of [that treatment leading to recovery]?"

Knowability. "Did the doctor have any way of knowing that this belief [about the probabilities] was false, or was it outside of her control?" (answer options: "Yes, there was evidence saying that her belief was incorrect" or "No, it was outside her control")

For Norming Studies S1 and S2, participants answered the following questions:

Choice. "What was the chance of success for the option which each person chose?"

Outcome. "Were the outcomes of these decisions generally good or bad?"

For Studies 1, 6, 7, and S3–S6, participants answered a subset of the following questions:

Belief. "Did the doctor know about the actual chances of success for each of the options?"

Choice. "What was the actual chance of success for the option which the doctor chose?"

Intention. "Did the doctor intend to choose the best treatment option for her patient?"

Knowability. "Was there existing evidence about the actual chances of success for each of the options?"

Specifically, participants in Studies 6 and 7 answered a generalized version of the belief question, a generalized version of the knowability question (Studies 6A and 7), and the choice question for each

vignette (Studies 6A and 6B). Participants in Studies 1 and S3–S6 answered the belief and choice questions, as well as either the intention question (Study S4 only) or knowability question (Studies S5 and S6).

Exclusion criteria. Because our hypotheses are predicated on the assumption that participants understand the agent's belief and choice, as well as the unknowability of the probabilities, participants incorrectly answering any of these questions were excluded from analysis for most of the studies (Studies 1–4 and S3–S5) and for the norming studies (S1–S2). However, the results generally did not depend on this decision. When the analyses were repeated on all participants for these studies (1–4 and S3–S5), the comparisons had similar significance levels as the primary analyses in nearly all cases (see Table S3).

For the remaining studies (Studies 5–7 and S6), we planned to look at individual differences and to compare the results across studies. Many participants in these studies misunderstood the "knowability" question asked in some of these studies. Therefore, we excluded any participant from these studies who incorrectly answered any question *except* the knowability question. Once again, this decision did not affect the outcomes of the analyses. When the analyses were repeated on only participants answering all questions correctly, all Best–Middle and Middle–Worst comparisons had similar significance levels.

Study	Best–Middle (*p*)		Middle–Worst (*p*)	
	Main Analysis	Alternative Criteria	Main Analysis	Alternative Criteria
1A	<.001	<.001	.78	.85
1B	<.001	<.001	.11	.30
2	.036	.20	.25	.098
3	.027	.052	.79	.088
4	.001	<.001	.67	.76
5	<.001	.018	.77	.37
6A	<.001	<.001	.066	.44
6B	<.001	<.001	<.001	<.001
7	<.001	.023	<.001	.001
S3A	.006	.001	.42	.66

S3B	.001	.001	.16	.24
S4A	.001	.001	.45	.026
S4B	.010	.002	.28	.15
S4C	.006	.001	.081	.12
S5A	<.001	<.001	.27	.69
S5B	.037	<.001	.48	.76
S6	<.001	<.001	.32	.37

Table S3. Response patterns in Study S6.

Note. Entries are *p*-values for the primary dependent measure in each study (blame and wrongness are averaged for Study 6), for the comparisons between the Best and Middle conditions and between the Middle and Worst conditions. For each pair of columns, these analyses are reported first using the exclusion criteria for the main text analyses and second using the alternative exclusion criteria described above.

References

Nobes, G., Panagiotaki, G., Pawson, C., 2009. The influence of negligence, intention, and out come on children's moral judgments. Journal of Experimental Child Psychology 104, 382–397.

Tversky, A., Shafir, E., 1992. The disjunction effect in choice under uncertainty. Psychological Science 3, 305–309.