Supporting Information:

Experiment 1:

Materials: Fifteen different shapes were used as the objects in Experiment 2. All fifteen shapes were roughly the same size and were constructed out of a plastic material using a 3-D printer. Each combination of shapes was counterbalanced between subjects to avoid the potential influence of perceptual features on participants' judgments of test objects.

Table S1: Descriptive statistics, Experiment 1

Age group	**Mean age (SD)**	**N (Number female)**
4 year olds	**4.88 (0.60)**	**74 (NA)**
6 year olds	**6.99 (0.67)**	**90 (NA)**
9-11 year olds	**10.27 (0.91)**	**90 (40)**
12-14 year olds	**13.11 (0.94)**	**86 (35)**
Adults	**NA**	**82 (NA)**

Table S2: Gender by condition, Experiment 1

	4yos	4yos_Fem	6yos	6yos_Fem	10yos	10yos_Fem	13yos	13yos_Fem	Adults	Adults_Fem
Conjunctive	25	NA	30	NA	30	14	28	12	28	NA
Baseline	24	NA	30	NA	30	16	29	12	26	NA
Disjunctive	25	NA	30	NA	30	10	29	11	28	NA

Table S3: Model comparisons for gender and experimenter effects
We compared statistical models of participant gender and experimenter gender in order to investigate the possibility of gender effects in our sample of older children and adolescents. Each model was compared to the null model and none of these comparisons reached statistical significance.

Models compared:	**Residual df**	**Residual deviance**	**df**	**Deviance**	**P**
M_1: Null model M_2: D judgment ~ participant gender	175 174	166.90 166.03	1	0.865	= .352
M_1: D judgment ~ Null model M_2: D judgment ~ experimenter gender	175 174	166.90 163.45	1	3.448	= .063
M_1: D judgment ~ Null Model M_2: D judgment ~ participant gender X experimenter gender	175 172	166.90 161.04	3	5.858	= .1187

Table S4: Model comparisons for 'D' test object

Models compared:	**Residual df**	**Residual deviance**	**df**	**Deviance**	**P**
M_1: Null model M_2: D judgment ~ Condition + Age group	421 415	486.32 358.84	6	127.48	< 0.001***
M_1: D judgment ~ Age group M_2: D judgment ~ Condition + Age group	417 415	438.30 358.84	2	79.46	< 0.001***
M_1: D judgment ~ Condition M_2: D judgment ~ Condition + Age group	419 415	417.78 358.84	4	58.95	< 0.001***
M_1: D judgment ~ Condition + Age group M_2: D judgment ~ Condition X Age group	415 407	358.84 349.10	8	9.74	0.284

*: Significant at $p < 0.05$ level
**: Significant at $p < 0.01$ level
***: Significant at $p < 0.001$ level

Table S5: Models comparisons for 'E' test object
Note that the degrees of freedom are 1 degree lower than the 'D' and 'F' object judgments, this is because one participant declined to make a judgment about the 'E' object.

Models compared:	Residual df	Residual deviance	df	Deviance	P
M_1: Null model	420	431.66			
M_2: D judgment ~ Condition + Age group	414	367.12	6	64.55	< 0.001***
M_1: D judgment ~ Age group	416	391.01			
M_2: D judgment ~ Condition + Age group	414	367.12	2	23.89	< 0.001***
M_1: D judgment ~ Condition	418	410.32			
M_2: D judgment ~ Condition + Age group	414	367.12	4	43.21	< 0.001***
M_1: D judgment ~ Condition + Age group	414	367.12			
M_2: D judgment ~ Condition X Age group	406	355.51	8	11.601	0.169

*: Significant at $p < 0.05$ level
**: Significant at $p < 0.01$ level
***: Significant at $p < 0.001$ level

Note: Results of McNemar's tests found that the new school-aged and adolescents participants in this experiment, like those in earlier studies and as predicted if participants used Bayesian inference, were significantly less likely to say that E ($M = 0.31$, $SE = 0.01$) was a blicket than D ($M = 0.48$, $SE = 0.01$; $p = 0.008$) in the conjunctive condition.

Table S6: Model comparisons for intervention choices

Models compared:	Residual df	Residual deviance	df	Deviance	P
M_1: Null model	421	486.32			
M_2: D judgment ~ Condition X Age group	407	367.18	14	125.21	< 0.001***
M_1: D judgment ~ Age group	417	484.37			
M_2: D judgment ~ Condition X Age group	407	367.18	10	117.19	< 0.001***
M_1: D judgment ~ Condition	419	398.72			
M_2: D judgment ~ Condition X Age group	407	367.18	12	31.53	0.001**
M_1: D judgment ~ Condition + Age group	415	388.63			
M_2: D judgment ~ Condition X Age group	407	367.18	8	21.44	0.006**

*: Significant at $p < 0.05$ level
**: Significant at $p < 0.01$ level
***: Significant at $p < 0.001$ level

Experiment 2:

Materials and procedure: The dolls used for 4 year olds, 6 year olds, 9-11 year olds and 12-14 year olds were named Sally, Josie and Mary. A trampoline, bicycle, and a diving-board were used as the toys that the dolls interacted with for these age ranges. Adults in the online version were presented with videos of dolls named Sally or Bobby interacting with a bicycle and a skateboard according to the condition. to which they were assigned. The order in which the dolls played, as well as the names of the dolls, and the types of toys they played with were counterbalanced between subjects.

If the children responded to the prompt for an explanation of the doll's behavior with an irrelevant answer (i.e. "I don't know") the experimenter gave them a forced choice question such as, "Do you think it's because the [toy] was [safe/dangerous] to play on or because [the doll] is the type of person who [gets scared/is brave]? Adults in the online version were simply asked to provide an open response answer to the question prompt: "Why did [the doll] [play/not play] on the [toy]?" Data from 34 adults were excluded from the analyses because they provided irrelevant explanations for the dolls' behavior.

Table S7: Descriptive statistics, Experiment 2

Age group	Mean age (SD)	N (Number female)
4 year olds	4.51 (0.31)	77 (33)
6 year olds	6.47 (0.37)	82 (39)
9-11 year olds	10.27 (0.91)	90 (41)
12-14 year olds	13.11 (0.94)	86 (35)
Adults	34.74 (11.3)	208 (89)

Table S8: Gender by condition, Experiment 2

	4yos	4yos_Fem	6yos	6yos_Fem	10yos	10yos_Fem	13yos	13yos_Fem	Adults	Adults_Fem
Person	24	11	24	11	30	15	29	13	73	40
Control	24	12	32	17	30	14	29	11	65	24
Situation	24	14	24	9	30	12	27	11	68	25

Table S9: Example responses to the explanation prompt by category

Person attributions:	Interaction attributions:	Situation attributions:	Irrelevant responses:
“She's confident in her bicycle riding ability.”	“She didn't get injured. It was available.”	"Because its [toy is] too high."	“I don’t know.”
"Because she might be more brave than the other one."	"Because its [toy is] her favorite color."	"Because her friend played."	“She was scared. The video does not explicitly say why she was scared, or I just did not get it.”
“She enjoys swimming.”	"It's fun and she's been on it before."	"Nothing was under the trampoline."	“The experimenter was making her move/not move.”
“Maybe she's afraid of heights.”	“Because it [toy] was maybe too high and she was catastrophizing.”	"The trampoline doesn't have any edges."	“There’s not enough information to tell.”
“She's afraid of trying new things because she had a bad experience of getting hurt.”	“Because it didn't have training wheels and she never rode a bike before.”	"It’s not safe"	“No clue”
"She's afraid of heights."	"Because she wanted to do what her friends were doing and the diving board was too shallow."	"Because its [toy is] too high."	No answer

Table S10: Number of participants (N) and explanation types in each age group by condition.

Condition (age group)	N	Person	Situation	Interaction	Forced choice
Person (4-year-olds)	24	37	8	3	15
Control (4-year-olds)	29	28	22	8	17
Situation (4-year-olds)	24	14	27	7	25
Person (6-year-olds)	24	42	7	1	2
Control (6-year-olds)	32	49	12	3	6
Situation (6-year-olds)	24	25	19	4	6
Person (9-11-year-olds)	30	51	4	5	14
Control (9-11year-olds)	30	40	13	7	8
Situation (9-11-year-olds)	30	24	33	3	8
Person (12-14-year-olds)	29	53	4	1	6
Control (12-14-year-olds)	29	50	6	2	3
Situation (12-14-year-olds)	28	17	27	0	4
Person (adults)	73	130	12	4	0
Control (adults)	65	100	20	10	0
Situation (adults)	68	75	46	15	0

Table S11: Model comparison information for gender effects for Experiment 2.

We compared statistical models of participant gender and experimenter gender in order to investigate the possibility of gender effects in our sample of older children and adolescents. Each model was compared to the null model and none of these comparisons reached statistical significance.

Models compared:	**Residual df**	**RSS**	**df**	**$SS_1 - SS_2$**	**F**	**P**
M_1: Null model M_2: Att. Score ~ participant gender	174 173	93.79 93.74	1	0.054	0.101	= .751
M_1: Att. Score ~ Null model M_2: Att. Score ~ experimenter gender	174 173	93.79 93.52	1	0.267	0.495	= .483
M_1: Att. Score ~ Null Model M_2: Att. Score ~ participant gender X experimenter gender	174 171	93.79 92.89	3	0.903	0.554	= .645

Table S12: Model comparison information for Experiment 2- Situation score by age and condition

Models compared:	**Residual df**	**RSS**	**df**	**$SS_1 - SS_2$**	**F**	**P**
M_1: Null model M_2: Att. Score ~ gender	539 538	284.73 284.19	1	0.541	1.024	= .312
M_1: Att. Score ~ Null model M_2: Att. Score ~ age group	539 535	284.73 271.96	4	12.771	6.281	< .001***
M_1: Att. Score ~ age group M_2: Att. Score ~ condition + age group	535 533	271.96 211.87	2	60.087	75.58	< .001***
M_1: Att. Score ~ condition + age group M_2: Att. Score ~ condition X age group	533 525	211.87 201.72	8	10.156	3.304	< .001***

*: Significant at $p < 0.05$ level
**: Significant at $p < 0.01$ level
***: Significant at $p < 0.001$ level