Gracia-Garrido, S. et al. (2021). Evaluating risk-taking in a cooperative context. Human Ethology, 36, 78-95. https://doi.org/10.22330/he/36/078-095

Human Ethology 36(2021): 78-95
Research Article

EVALUATING RISK-TAKING IN A COOPERATIVE CONTEXT

Santiago Gracia-Garrido[1], Marcos F. Rosetti[1,2], Robyn Hudson[1]
[1]Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico
[2]Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz, Mexico City, Mexico

mrosetti@gmail.com

ABSTRACT

Risk-taking is a fundamental feature of human behavior. The evaluation of risk-taking has commonly focused on the assessment of individuals rather than on social scenarios which often entail cooperative interactions as well as consensual decision making. Our aim in the present study was to assess joint risk-taking in same and mixed gender dyads from an age range encompassing children and young adults. For this, we tested participants in an ecologically relevant, cooperative tower-building task in which they had to work together using their assigned wooden blocks to build the tallest tower they could within 10 minutes. Participants of all ages collaboratively engaged in a construction process that involved options, uncertainty, and a potential for undesirable outcomes. We found that adult male dyads built taller towers than female and mixed dyads. Given the low number of metrics showing gender differences and the small effect sizes, we consider further methodological implementations in order to increase the salience of the outcome and in consequence, provide a sharper assessment of risk-taking. We conclude that the current task shows potential as a novel experimental method to evaluate risk-taking in a realistic cooperative context.

Keywords: *cooperation, dyadic interactions, risk-taking*

---

submitted: 26.Jan. 2021; revised: 19. May 2021; accepted: 05. July 2021

INTRODUCTION

Decisions are commonplace in our daily lives, and they are accompanied by uncertainty as to their outcome, whether that be positive or negative. At the center of such decisions lies risk-taking, which involves a potential for loss while simultaneously providing an opportunity to obtain some form of benefit (Leigh, 1999). While mathematical theories exist about which decisions are the most favorable (Von Neumann & Morgenstern, 1947), the variability and whim of human decision making processes has not conformed to these expectations (Kahneman & Tversky, 1984); it is this contrast between what is theoretically expected and what happens in our daily decision making that makes this topic particularly fascinating.

Risk-taking is a ubiquitous aspect of human behavior. From daily and harmless to rare and life-changing decisions, risky behavior permeates a large number of life scenarios: telling a joke, gambling, driving, unprotected sex or drug consumption (see meta-analysis by Byrnes, Miller, & Schafer, 1999). Even daily activities such as crossing a busy road or adjusting arrival time at a bus stop to cut waiting involve a degree of risk (Pawlowski, Atwal, & Dunbar, 2008). Although the decisions concerning risk-taking vary among such contexts, three basic elements are thought to be common to all: options, outcomes, and uncertainties (Fischhoff & Kadvany, 2011).

Risky behaviors have often been evaluated in relation to gender and age, finding that men showed greater risk-taking than women (Byrnes et al. 1999; Cobey, Laan, Stulp, Buunk, & Pollet, 2013; Pawlowski et al., 2008; but see Nelson 2015 for a perspective contesting this) and that this difference between genders is present from a young age (Amir et al. 2020; Ginsburg & Miller, 1982; Slovic, 1966). In terms of the stability of risk-taking across development, Byrnes et al.'s (1999) metanalysis of 150 studies reports significant changes in the size of the gender gap between successive age groups (particularly among self-reported, prototypical risk-taking behavior such as drinking, sex and drug use, but also among observations of physical activity in the case of younger age groups), while also remarking that the gender gap reduces as individuals advance in age.

Another important aspect of risk propensity pertains to the effect of the social environment. Mostly, the relationship between risk-taking and the social environment has been addressed by research using economic games (Suijs, 2012) and social dilemmas (Kirley & von der Osten, 2014). Some examples of interest are cooperation among spouses (Cochard, Couprie, & Hopfensitz, 2016), competitive vs cooperative dyads (Lupfer, Jones, Spaulding, & Archer, 1971), or individuals vs dyads (Deck, Lee, Reyes, & Rosen, 2012). While certainly very useful, these types of tasks greatly simplify cooperative interactions, which in the real world often involve extensive interaction between agents. Examples of scenarios where decisions are usually discussed and agreed on mutually include going on a trip, raising a child, discussing each partner's financial contribution to the mortgage on a house, among others. These are all situations that involve joint risk-taking as there is an investment (time, money or emotional strain) and potential danger or loss (cost) as well as a cooperative effort (i.e., the pursued goal is most likely to be successfully achieved by the combined behavior of all members in the group; Keller & Schoenfeld, 1950). Such situations often involve seeking consensus among those concerned. In social contexts, discussion within groups could lead to different outcomes than to those of individuals acting alone. Rao et al. (2016) found that it is more likely for women to choose the safer option and undergo childbirth at a healthcare facility when the decision was made jointly with partners. In contrast, when comparing single individuals and same-gender triads, Gardner and Steinberg (2005) found the latter showed greater risk propensity, suggesting an effect of peer influence. Booth and Nolen (2012) found that girls in single-gender groups were more likely to choose a real-stakes gamble than those in mixed-gender groups. Furthermore, Stöckl, Huber, Kirchler and Lindner (2015) found that the gender composition of the dyad had a modulatory effect on investment behavior, with all-female dyads choosing the risk-free investment strategy significantly more often than all-male and mixed-gender dyads. While these studies show that risk propensity can be influenced by the presence of a partner as well as by the gender composition of the group, an equally successful solution could be achieved either individually or cooperatively. The notion of cooperation that the present study aims to address entails the interaction of two individuals performing interdependent roles to support each other's action towards a common target (Warneken & Tomasello, 2007) or as Bratman (1992) puts it, requiring the ability to mesh plans of action toward a shared goal.

Risk-taking has been addressed from a wide range of disciplines, with methods ranging from self-report (e.g., Pfefferbaum & Wood, 1994) and self-administered scales (e.g., Zuckerman, 2007; Zuckerman, Kolin, Price, & Zoob, 1964), hypothetical financial scenarios (e.g., Gneezy & Potters, 1997; Kahneman & Tversky, 1984), inferences drawn from datasets relating to antisocial behavior and other factors associated with a gender disparity in mortality rates (Kruger & Nesse, 2004; Moffitt, Caspi, Rutter, & Silva, 2001), or from observational studies (e.g., Pawlowski et al., 2008). Such a broad scope remains one of the principal challenges in the assessment of risk preferences. In order to aggregate behaviors from different contexts into general risk-taking profiles, authors have developed scales which group risk-taking activities into "domains" (Blais & Weber, 2006; Wilke et al. 2014). The focus on a single "domain" is also evident from behavioral assessment methods of risk-taking. The many existing methods, such as the Iowa Gambling Task (Bechara, Damasio, Damasio, & Anderson, 1994), the Balloon Analogue Risk Task (Lejuez et al., 2002) and the Columbia Card Task (Figner, Mackinlay, Wilkening, & Weber, 2009) principally assess a "financial" domain, as they all represent different embodiments of gambling tasks. Furthermore, these methods (i) almost exclusively evaluate participants individually (but see Fischer & Hills, 2012), (ii) remove the influence of motor skills by implementing scenarios where decisions are carried out by pressing a button or drawing a card, (iii) seek to standardize motivation by providing monetary or edible rewards and (iv) explicitly involve risk-taking, that is, the role of risk is an essential part of the instructions received by the participant. While such standardization facilitates comparisons across samples, it also limits risk assessments to a single domain.

These methods illustrate how diverse risk assessment can be, and yet, to our knowledge none is specifically designed to address a cooperative interaction involving joint risk-taking. Here, we propose a task based on a different approach. For this, we developed a cooperative task that allows for the evaluation of joint risk-taking. We believe the task is ecologically relevant, in that it attempts to incorporate the interactive aspects of real life joint risk-taking scenarios. The Tower Building Task, or TBT, involves using wooden blocks of uniform size to try to build the tallest tower possible within a limited time. To do this, a pair of participants needs to work together, each member using only blocks of the color assigned to each of them. The number of blocks assigned to each is the same but limited, and thus a joint effort by participants is needed to ensure a favorable outcome. In this task motivation is intrinsic; as with many tabletop games, fulfilling the aim of the task and attempting to "win" by building a tall tower is expected to provide its own reward. Like in many daily scenarios, the aspects of risk-taking are up to the participants to judge. They decide how high to build the towers (according to their own perceived risk, judging their own level of skill, etc); similarly, what counts as a "loss" (for instance, in the event of a collapse) is up to them. As in other situations involving risk-taking, sensory feedback can allow participants to judge the stability of the structure and decide whether or not to continue building. Since the current version of the TBT allows for multiple attempts to construct towers, participants can learn from first attempts and change their strategy in further attempts.

Tower-building paradigms have been previously used in the assessment of human behavior. For instance, a task involving the construction of towers using clay and spaghetti by chains of participants was used to simulate the processes of cumulative cultural evolution (Caldwell & Millen, 2008; Reindl, Apperly, Beck, & Tennie, 2017), whereas a tower-building task involving interlocking plastic cubes and flat squares was used to measure innovation following observation (Subiaul & Stanton, 2020). Finally, Jenga tower collapses have been used for didactic purposes as metaphors for the interdependence and fragility of ecological systems (Umphlett, Brosius, Laungani, Rousseau, & Leslie-Pelecky, 2009).

Our aim in the present study was to assess risk-taking in a cooperative context by means of the TBT. For this, we tested same- (all-male, all-female) and mixed-gender (male-female) dyads within a wide age range (three age groups: six-, 12-, and 18-year-olds). In accordance with previous studies, we expected (i) to find gender differences in each of the age groups, with all-male dyads engaging more frequently in risky behaviors than all-female dyads, and (ii) for this to be modified in mixed-gender dyads, resulting in scores intermediate to those of same-gender dyads. Additionally, we (iii) expected any such gender differences to be apparent from an early age. We also aimed to (iv) use the current experimental design to establish a performance baseline for the most basic form of the cooperative TBT.

METHODS

Participants

We tested 165 dyads (330 participants) from three age-groups: six-year-old (6YO, mean age = 6.28, sd = 0.29) and twelve-year-old children (12YO, mean age = 11.6, sd = 0.358) from three public elementary schools in Mexico City, and university students (18 mean age = 19.3, sd= 1.34) were recruited on the main campus of a public university, also in Mexico City. Age groups were roughly 6 years apart, ages that mark important developmental milestones (start of academic life, start of adolescence, and start of adulthood, respectively). We allocated participants within each age group to dyads according to one of three gender combinations: same- (female [F] and male [M]) and mixed-gender dyads (MIX). Participants' gender was not asked but inferred from their names in the class registers. Participants were only tested once. Dyad members, including university participants, were classmates, but were paired randomly by the experimenter based on their number in the class register which was obtained beforehand. Twelve dyads were omitted after testing: Six of them due to failure of recording equipment and another six (two of each gender combination, all belonging to the 6YO group) because they failed to cooperate, i.e., not building a single tower together during the whole trial despite corrective comments from the experimenter. This resulted in 17 dyads per condition which were included in the analysis.

Tower Building Task

The task consisted in having a single dyad build together the tallest tower they could using wooden blocks (1.5 x 2.5 x 7.5 cm, Fig. 1a) from the board game *Jenga* (Parker Brothers™, Hasbro Inc, USA). In the current task, each member of the dyad received 54 blocks of a single color, either red or blue (Fig. 1b). To ensure the cooperative nature of the task, each participant was asked to only handle blocks of the color assigned to him or her. To eliminate the effect of small variations in floor topography, we instructed participants to build on a 50 x 50 cm board with a smooth melamine surface (Fig. 1c). A large hourglass was placed conspicuously beside the board to inform participants of the time remaining for the task (Fig. 1d). Dyads were also told they could keep attempting to build their tower until time on the hourglass had run out (10 minutes) or stop at any moment if satisfied with their result.

Figure 1: Experimental setup. Dyads were instructed to build the tallest tower they could (a) using only blocks from the pile they were assigned (b) over a flat, uniform melamine surface (c) before time on the hourglass ran out (d).

Dyads received the following instructions: "*(i) The goal is to build together the tallest tower you can within 10 minutes; it is a team effort. (ii) You can only use the blocks of the color that was assigned to you. (iii) If the tower collapses you can keep building until time is up. (iv) If satisfied with the tower, you can stop building before the time finishes.*" If participants asked questions about the building procedure (e.g., "*Can we put two blocks at the same time?*" or "*Can I talk to him/her?*") these were answered with standardized responses ("*You can build in any way you want*" and "*Yes, talking to each other is allowed*"). When participants infringed basic rules of the task, such as taking blocks from their partner or building a structure outside the melamine board, these were corrected with a brief verbal reminder from the experimenter. When testing the youngest dyads, we gave slower instructions using simpler terms, and ensured that participants understood the instructions by asking them questions and having them explain the task in their own words. We also took care to explain the purpose and to ask about the function of the hourglass, which all participants seemed to understand.

Procedure

Using the class register, participants' names were called, and each dyad accompanied the experimenter to the test area where instructions were given. The task was performed by one dyad at a time and out of sight of other participants. Trials were conducted either in an empty classroom or an unoccupied playground close to where the participants were recruited. Testing was conducted by the same male experimenter during school hours (09:00 to 14:00 hrs), and before the test, members of each dyad consented to participate in the study and were asked their name and age. Instructions were then given, and the two participants took their place beside their assigned set of wooden blocks. Trials were filmed for later behavioral analysis using a video-camera (Handycam CX405, Sony Corporation, Tokyo, Japan) installed on a tripod about 3 meters from the participants. A whiteboard giving the experimental condition (F, M, MIX), and the date, time of day, and age group of the participants, was filmed for a few seconds before telling them to start the task. The experimenter (SG) was present during all trials to control the camera and correct participants' behavior when necessary.

Recruitment and experimental procedures met the bioethical requirements established by the Internal Review Board for Research with Human Subjects of the Instituto de Investigaciones Biomédicas, UNAM.

Behavioral coding

Video files of all trials were analyzed using event logging software (Friard & Gamba, 2016), in which we coded the sequence of block additions, tower collapses that occurred during each trial, and early conclusion when these took place. These events were later transcribed into behaviors considered indicative of participants' performance and risk-taking propensity (Table 1). Other aspects of participants' behavior indirectly related to risk-taking were also analyzed (for a description of these measures, see part 1 of Supplementary material: Additional behavioral analysis).

We recruited a second rater trained in scoring the behaviors of interest to register the sequence of events from two randomly selected videos from each category (n = 18, 11.7%). Subsequently, we calculated interrater reliability by comparing the metrics obtained from each rater by means of a Pearson correlation for continuous metrics. All values were significant and equal to or above $r > .94$. Total collapses were tested using a Spearman correlation and had an $r_s = 1$.

Table 1: Behavioral descriptions

Type	Dependent variable	Description	Explanation
Building	Maximum height	Single tallest tower achieved during the trial.	In general, taller towers have a larger probability to collapse; larger heights may be indicative of riskier choices.
	Height gain	Height gained (cm) per piece added.	Height gain may be understood as the payoff at the expense of towers' stability; higher scores may suggest a higher risk propensity.
	Addition rate	Number of pieces added per second. All pieces added were considered regardless of their placement on the board or as part of the structure.	The speed with which pieces are placed can provide information about how carefully pieces were placed.
	Proportion of vertical pieces	Number of vertical pieces divided by the total number of pieces used throughout the entire building process.	Since vertical positions result in a higher payoff and reduced stability; therefore, they may be considered as a risky choice.
Collapses	Involuntary	Unwanted loss of tower height meeting the following criteria: (i) a loss of at least a quarter of the height constructed, (ii) at least a quarter of the pieces that make up the tower must fall and (iii) the tower had to consist of at least 10 pieces.	An involuntary collapse is not a risky behavior, but rather the negative consequence of the accumulation of risky choices leading to the loss of height.
	Demolitions	Occurred when dyad members, presumably unsatisfied with their performance, voluntarily collapsed the tower (either partially, by disassembling a part of it, or totally) to improve on their previous attempt. Same criteria as mentioned above were applied.	Demolition implies another attempt by discarding the built tower, thus it entails the implementation of an uncertain but potentially better option.
	Clumsiness	Resulted from events such as accidentally bumping the board or involuntarily knocking the tower with any part of the body.	Incidents of clumsiness were not necessarily related to risk, and thus, these collapses were not included in the analysis.
Time	Latency to start	Time elapsing between the start of the test and the time when the first block was placed.	Longer latencies could reflect verbal interaction leading to planning, rehearsing and reaching consensus.
	Early conclusion	The time elapsing between the moment participants placed the first block to the time they stopped building. Only decisions to stop construction taken 100 seconds before the allotted time, which was sufficient time to attempt a new tower, were considered.	An early conclusion suggests an unwillingness to engage in a new, uncertain attempt, and thus it can be regarded as a form of risk aversion.

Statistical analysis

Statistical analysis was done using R (R Core Team, 2020). We chose to use the metrics shown in Table 1 as dependent variables as they reflect different aspects of risk-taking. We used one-way analyses of variance (ANOVAs) to compare the behavior among gender combinations in each age group. For the ANOVAs we calculated the effect size (η^2) using the function *eta_sq* from the package sjstats (Lüdecke, 2018). In the case of significance, we also performed *post hoc* Tukey's Honest Significance Tests corrected for multiple comparisons and report the confidence coefficient for the set.

We also assessed changes in the construction performance of participants after experiencing a collapse. This analysis only included dyads who built more than one tower (n = 104, 68%). For this, we used general linear mixed models to evaluate the effect of gender composition within each age group and the number of attempts on the addition rate, height, and height gain for up to 5 consecutive towers. In these models, data were grouped by the number of towers and the dyads' id, fitted as random intercepts. In all cases, diagnostic plots showed symmetrical residuals with a heteroscedastic distribution, with a bias towards lower values.

All plots were done using the package ggplot2 (Wickham, 2009). All tests were two tailed and significance was set to $p < 0.05$.

RESULTS

General description of tower-building behavior

All groups readily engaged in the construction of a tower, which varied considerably in height among dyads (mean = 54.98, sd = 27.05, range = 9 – 171 cm; Fig. 2a) and clearly showed that blocks were added, at least for the two older groups, with the purpose of increasing tower height (Fig. 2b). Towers collapsed rather often, with 68% of trials having at least one collapse (Fig. 2c; for more information, see part 2 of Supplementary material: Breakdown of collapse types), yet participants continued building, with a mean of 2.34 (sd = 1.38) towers per trial, including the first attempt.

Participants started the task quite quickly (mean = 12.69 s, sd = 9.16) and when a collapse occurred, they rapidly went back to building (mean = 16.82 s, sd = 19.98; for more information, see part 3 of Supplementary material: Latencies). Less than a third (n = 45, 29%) of the dyads chose to end the task before the allotted time, usually rather late in the test (on average after 346.22 s, sd = 102.50, that is, when at least 58% of the allotted time had elapsed; Fig. 2d). From those dyads that chose an early conclusion, 10 were 6YOs (3 females; 2 mixed; 5 males), 22 were 12YOs (8 females; 8 mixed; 6 males) and 13 were 18YOs (4 females; 5 mixed; 4 males).

Dyads added blocks at a mean rate of one block every 2.98 seconds (sd = 1.01; Fig. 2e). Participation by both members of the dyad was evident from visual inspection of the construction sequences, which showed clear color alternations, even when turn taking did not always involve one block per turn (for more information, see part 4 of Supplementary material: Examples set of construction sequences).

In general, building structures were diverse despite only a mean 8% of the pieces being placed vertically (Fig. 2f), and included structures such as arches (a horizontal piece over two vertical pieces), antennas (two or three vertical pieces stacked on top of each other placed on top of the tower), or alternating pairs of parallel horizontal pieces leaving a space in the middle. A sample of building arrangements is provided in part 5 of Supplementary material: Examples of building arrangements.

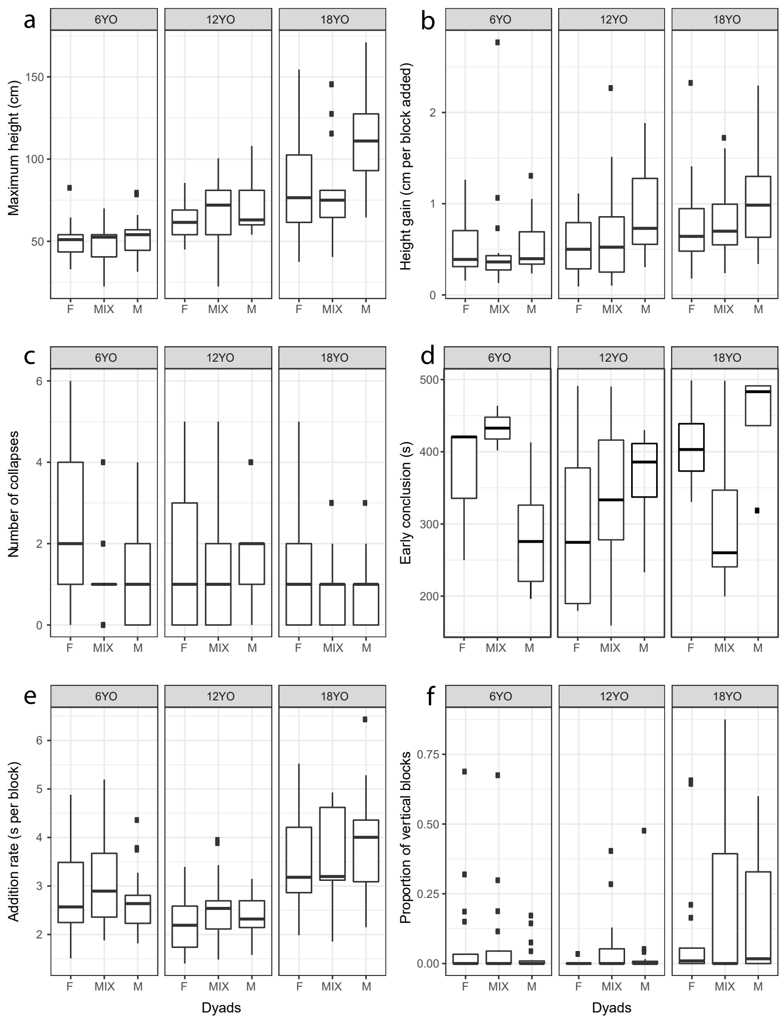

Figure 2

Figure 2: Construction performance. Boxplots for each age group and gender combination of the (a) maximum height (cm), (b) height gain (cm gained per block added), (c) number of collapses, (d) early conclusion (s), (e) addition rate (s per block) and (f) proportion of vertical blocks. Horizontal lines through the boxes mark median values, box limits represent the 1st and 3rd quartiles. Whiskers extend from the limits of the boxes to the smallest and largest values no further than 1.5 times the interquartile range. Outliers (filled circles) lay beyond this range.

Tower-building and risk-taking

When comparing performance among the different gender combinations within each age category, we found that the 6YO female dyads had significantly more collapses than the 6YO mixed dyads and that 18YO all-male dyads built taller towers than the 18YO mixed and all-female dyads (Table 2).

Table 2: Significant results of analyses of variance comparing gender combinations in each age group and the corresponding *post hoc* comparisons.

Metric	Analysis of variance	Post hoc comparison	Estimate	Confidence interval	Adjusted *p*-value
Maximum height	$F = 4.33$, $df = 2$, $p = .02$, $\eta^2 = .15$	18YO: F vs M 18YO: MIX vs M	-26.62 -25.32	-1.95 -- -51.29 -.66 -- -49.99	.03 .04
Total number of collapses	$F = 3.48$, $df = 2$, $p = .04$, $\eta^2 = .13$	6YO: F vs MIX	1.35	-2.62 -- .08	.03

In relation to the behavioral descriptions of Table 1, we could interpret that 6YO all-female dyads suffered the consequence of the accumulation of risky choices, although a lack of significant differences in tower height between gender combinations in this group may suggest that this may be related to motor skills, rather than risk propensity. As for the significantly taller towers built by the 18YO all-male dyads, we would interpret this as the result of greater risk-taking, as the lack of differences in collapses, in addition rate or in proportion of vertical pieces among the other gender combinations at this age would suggest that they were similarly skillful, and that the all-male dyads decided to try to reach greater heights (see Discussion for an explanation of how this may have taken place). This difference was not found in the other age groups. Additionally, after the significantly taller towers of 18YO all-male dyads, the lack of significance in terms of height gain was unexpected. We attribute this to the fact that towers, in general, contained very few vertical pieces and the height gain achieved by such additions was diluted by the large number of horizontal pieces or several vertical pieces at the same level. No significant differences were found for the other comparisons. For more details, see part 6 of Supplementary material: Examples of *post hoc* comparisons.

Also, we found that when building subsequent towers, height, height gain and addition rate did not substantially change as the number of towers increased or when comparing gender combinations. That is, the basic pattern did not change between groups and 18YO male dyads continued to build higher towers across subsequent attempts (Figure 3, Table 3). Note, however, that the decreasing slopes generally observed on all age groups and gender combinations may reflect that latter towers were shorter because participants ran out of time. For more details, see part 6 of Supplementary material: Examples of *post hoc* comparisons.

Table 3: Significant results from the general mix effect models evaluating the effect of consecutive towers and gender combination on selected behavioral descriptors, summarized in an analysis of variance showing the contribution of each term and the corresponding *post hoc* comparisons.

Metric	Term	Analysis of variance	Random effect*	Pos hoc comparisons	DF	Estimate ± St. Error	Adj *p*-value
Height gain (cm per piece added)	N. of towers	$F(1, 40)$, p = 0.75	0.12	18YO: F vs M	26	-0.53 ± 0.18	0.02
	Dyad	$F(2, 26)$, p = 0.01		18YO: MIX vs M		-0.56 ± 0.19	0.02
Height (cm)	N. of towers	$F(1, 40)$, p = 0.08	32.48	18YO: F vs M	26	-33.82 ± 9.95	< 0.001
	Dyad	$F(2, 26)$, p < 0.001		.18YO: MIX vs M		-38.66 ± 10.15	< 0.001

Note: *Random factor for the grouping *N. of towers* and *Dyad's id*

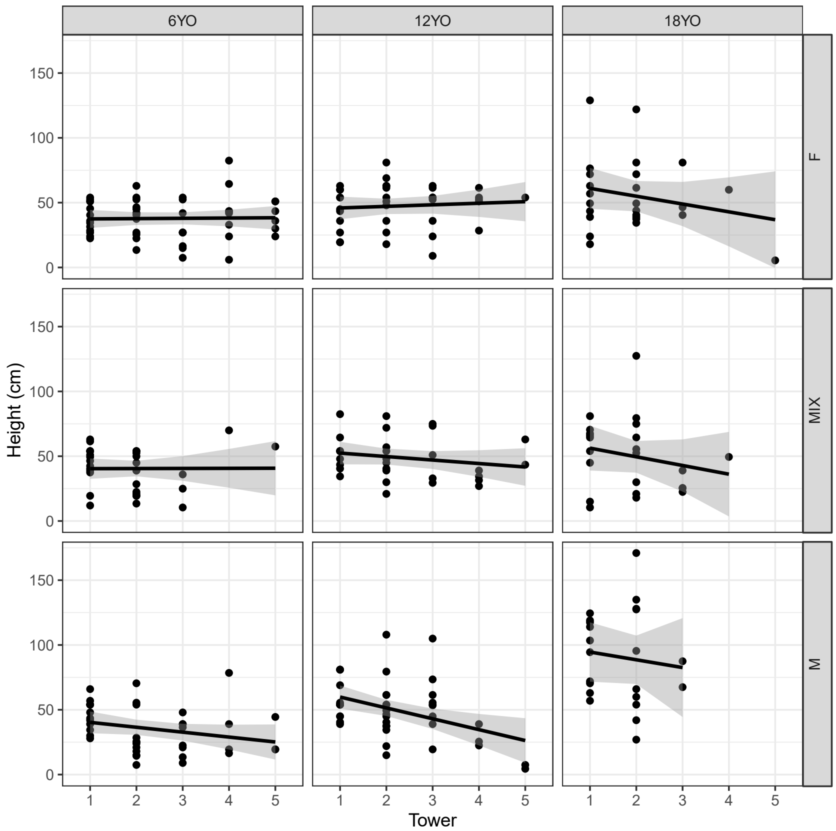

Figure 3: Sequential construction events. Scatter plots displaying the height of sequential towers built by each age group and gender combination. Filled circles show the height of each individual tower. Black lines show the linear regression while gray shading marks the confidence intervals.

DISCUSSION

Our aim in the present study was to assess cooperative risk-taking using the TBT in same and mixed-gender dyads of an age range encompassing children and young adults. We looked for differences between gender combinations within each age group. We found that all 18YO all-male dyads built taller towers than the other gender combinations despite similar building efforts, possibly suggesting a larger degree of risk-taking on their part. Below, we discuss each of the findings and their limitations, as well as methodological considerations regarding the TBT.

Risk-taking and gender combination

We found that all-male university student dyads built taller towers than all-female and mixed dyads. However, these same dyads did not differ in terms of height gain per piece added, addition rate, proportion of vertical pieces, number of collapses and early conclusions. Not only did the 18YO all-male dyads construct overall higher towers, they continued to build higher towers after collapses, suggesting a reiterated commitment to riskier choices.

Initially, we might interpret the difference in the maximum height of the towers as suggesting a gender disparity in terms of risk propensity. The maximum height of the towers stems from building arrangements, which despite containing a similar number and orientation of blocks (see addition rate and proportion of vertical pieces), resulted in different maximum heights and less contact between the blocks (i.e. less stability). For instance, an arch composed of two vertical pieces and a horizontal one would result in a height of 9 cm (with a height gain of 3 cm per piece), while an antenna composed of a horizontal base and two vertical pieces, one atop the other, has the same number of components but adds 16.5 cm (with a height gain of 5.5 cm per piece). Clearly the latter entails a larger payoff with less stability (i.e. is a riskier option). Current results suggest males opted more often for these kinds of choices. Nevertheless, following the caveats of Nelson (2015) and Boyer and Byrnes (2016), we acknowledge that this difference singles out one metric and was accompanied by low effect sizes.

As with many real-life activities, contributing factors such as skill and motivation have been traditionally difficult to disentangle within some risk domains (e.g., the recreational domain, which includes games, sports, and outdoor activities). Regarding the potential confounds of cognitive and motor skill in relation to tower height, research concerning fine motor movements in young adults found no evidence of differential performance between genders (Peters, Servos, & Day, 1990). Whereas widely used tests of visuospatial abilities have shown that males outperform females with regard to this type of skill (Shepard & Metzler, 1971; Vandenberg & Kuse, 1978), when applying new, more comprehensive approaches, such differences were not found (Fisher, Meredith, & Gray, 2018). A more difficult aspect to isolate is motivation. Differential motivation can result from the well-documented gender divide in preferences for particular types of toys and activities; e.g., a larger number of boys than girls reportedly show a preference for manipulable toys, such as those involving building and design (Cherney & London, 2006), which might lead males to be more prone to build a taller tower. Differential motivation could also arise from the cultural expectation that males should develop a more competitive attitude, as shown by studies where males from patriarchal societies and/or western cultures show a stronger tendency for competitiveness (Andersen, Ertac, Gneezy, List, & Maximiano, 2013; Gneezy, Leonard, & List, 2009), as “build the tallest tower” carries an implicit competitive connotation. Linked to this, the relatively shorter towers of female dyads could result from them underestimating their building capabilities, as has been reported to occur when females are faced with "masculine" tasks (Beyer, 1990).

Finally, the inclusion of mixed-gender dyads in the experimental design was based on the expectation of intermediate scores for this group. While we did not find intermediate scores for the mixed group, modulation could be seen in the similar maximum height of the towers of the all-female and mixed 18YO dyads, with both being significantly smaller than those of 18YO all-male dyads. There is little information to date with which to compare such findings, as only a few studies have focused on the behavior of mixed-gender groups and particularly on the way in which one gender might modify its behavior in the presence of the other in a context of risk-taking (e.g., Booth & Nolen 2012; Stöckl et al., 2015). If we interpret tower height as a proxy for risk-taking, then the current results contrast with those of Stöckl et al. (2015), who found mixed dyads' risk propensity did not differ from that of all-male dyads.

Risk-taking and age

We did not find differences when comparing dyad combinations in the 6 and 12YO age groups.

While the current methodology appears to be informative regarding risk-taking inclinations by adults, it is unclear whether risk-taking assessment can be carried out using this task in the younger dyads. Both, 6 and 12YOs built towers cooperatively at least at some point of the task (evidence of joint effort and cooperative interactions come from alternation in the placement of blocks of the two colors in most constructions) while implementing a variety of options; for instance placing blocks vertically (albeit very few) or choosing to continue building regardless of the tower's instability (as reflected by numerous collapses).

A first concern pertains to the 6YO children's building procedure. Most participants in this group built relatively flat towers with a low height gain, starting rapidly, and hurrying to use up all the given blocks within the time limit. In such cases one could conclude that options, potential benefits and negative outcomes were not necessarily linked to vertical arrangements of pieces or towers collapsing, but rather to use all the pieces or to finish before the time was up. So, if this was the case, what were they risking and what was the valued outcome for them? Currently, we cannot answer this question. The behavior of the 6YOs contrasts with the findings of Slovic (1966) and Ginsburg and Miller (1982), who reported a capacity for children to engage in risky behaviors from early childhood while understanding the consequences. A crucial difference between these studies and our own is the evaluation of joint risk-taking. In this sense, one of the reasons risk-taking may not be evident in our sample is that the assignment of an unfamiliar peer may have created an awkward situation (Hartup, 2009), hampering their coordinated action and promoting the large number of collapses observed in the 6YO group. These aspects could be further explored. One option could be to utilize aspects of the tower building tasks shown in Reindl et al. (2017) or Subiaul & Stanton (2020). Allowing children to observe others or even older participants could help homogenize building modalities before the evaluation takes place and might help younger participants appreciate the benefits and consequences of different building strategies. These concerns suggest that the TBT in its present form may not be sufficiently sensitive to assess cooperative risk-taking at early ages. In the following section we address modifications that could make the task more appropriate for this, and indeed also for older age groups.

We were surprised by the lack of gender differences in the 12YO group, as reportedly the levels of rewarding stimulation associated with novelty and sensation seeking increase dramatically at puberty (Steinberg, 2004), especially among males (Cross, Cyrenne, &

Brown, 2013). One consideration is that risk-taking at this age is strongly affected by the influence of peers (Jessor & Jessor, 1977). Since we randomly paired dyads, the social situation we created may not have been the one needed to observe this kind of behavior. Furthermore, it may even have precipitated the opposite, with participants choosing to behave conservatively in order not to damage their social status by performing poorly (e.g. by precipitating collapses). A second consideration is that participants in this age group were not engaged enough by the task in its current form. Since the risk domains that seem to take over during adolescence are those related to novelty and sensation seeking (Kelley, Schochet, & Landry, 2004), perhaps the stakes involved in the tower-building task need to be higher (competition against another pair, break a previously set record, limit the number of collapses) for participants of this age to engage effectively in such a task.

On the Tower-Building Task

Our final aim was to establish a baseline for performance on the TBT. Despite certain limitations, we think the TBT can contribute in several ways to the study of cooperative risk-taking behavior, particularly to scenarios in the recreational domain.

In cooperative economic games, social interaction is often reduced to a single decision (i.e. cooperate or not) and a single event. In contrast, the interaction between participants on the TBT is closer to real-life cooperative situations. In the TBT cooperation can be graded (i.e. cooperate, not cooperate or partially cooperate) and cooperation takes place over a relatively long series of events. A task like the TBT could serve to evaluate these dynamics by looking in detail at interactions taking place during the construction process, e.g., discussing a particular structure, copying a structure built by the other participant, switching strategies after the one by the other partner not being very effective, etc. In addition, most risk-taking assessment tasks are screen-based (e.g., Bechara, et al., 1994; Figner, et al., 2009; Lejuez et al., 2002), a medium that makes it difficult to capture the level of nuanced interaction and variability in execution that a task like the TBT can offer (c.f. Rosetti et al. 2017). It would be difficult to capture the nuances of, for instance, a piece put in a precarious position or the social gestures involved in the interaction by virtual means. Lastly, we suggest that differences between the TBT and other standardized risk-taking tasks could enrich the range of presently available means of assessment. The TBT, for instance, could serve to assess behavior in a different domain (i.e. recreational) since it contains a physically realistic negative outcome that affects both participants simultaneously (the tower collapsing) but that is nevertheless without harmful consequences. In spite of the physically harmless nature of the TBT, it fulfills the basic criteria of an operationally broad definition of risk consistent with its three fundamental elements, or as Byrnes et al. (1999, p. 367) succinctly put it "*…the implementation of options that could lead to negative consequences*".

Nevertheless, there are several aspects that could be improved in future testing, particularly pertaining to the experimental design and robustness of the resulting behavioral descriptors. Initially, we considered involuntary collapses as one of the most informative risk indicators as these are conspicuous and integrate the sequence of choices into a single event. However, during the analysis we noticed that rather than risk, at younger ages these were probably linked to immature motor and social skills; the number of collapses was largest for the 6YO dyads despite this group building the lowest towers. Collapses, however, could become a relevant descriptor by increasing the salience of the outcome. This could be done by encouraging conservative behavior (e.g., by limiting the trial to a single collapse), or conversely, by encouraging participants to take risks, for instance, by explicitly stating that they are competing against other dyads, or providing a reference value (e.g., a previous spurious "record" to beat). If trials were to end after the first collapse, a decision by participants to conclude the trial early could provide a more robust risk aversion indicator: the earlier that dyads opt to conclude the trial, the lower their propensity for risk-taking. Indeed, limiting the test to a single tower and collapse may be sufficient given that participants did not seem to shift their strategies with repeated attempts.

Whether the building process came to a halt because participants were satisfied with the tower height, ran out of blocks, or both, the choice to stop building despite having sufficient time to rebuild could represent a risk-averse choice. Concerning the availability of building material, when all blocks were used up, dyads still motivated to increase tower height could either (i) demolish the current tower so as to engage in a new attempt or (ii) make deletions, that is, taking a block(s) from the current structure with the purpose of adjusting the construction to achieve greater height. Demolitions are arguably the opposite of concluding the task early, as they entail the implementation of an uncertain option for a potentially larger payoff in a subsequent attempt (i.e., risky option). Deletions were rarely performed, probably because they were perceived as a poor choice, involving the possibility of causing a collapse while achieving only a minor gain in height.

In conclusion, we found behavioral differences on the TBT among young adults, suggestive of a disparity in risk propensity related to gender composition. However, the task in its current form, was not ssufficiently sensitive to evaluate risk-taking at an early age. Nevertheless, the task provides a first step towards the evaluation of risk-taking in cooperative scenarios, as it involves coordinated efforts to achieve a shared goal, which entails options, different outcomes and uncertainty. Additionally, the task could be particularly useful in the evaluation of behaviors within a certain risk domain, such as the recreational one. We consider that the baseline provided by the present study can be used as a cornerstone to develop further versions of the TBT. Such versions should involve more costly decisions so as to make risk preferences more evident and in turn, contribute to a better understanding of the interactions that take place in risky but cooperative contexts.

ETHICAL STATEMENT

These studies were approved by the ethics committee for research with human subjects of the Instituto de Investigaciones Biomédicas of the Universidad Nacional Autónoma de México.

CONFLICT OF INTERESTS

On behalf of all authors, the corresponding author states that there is no conflict of interest.

DATA SETS

Construction sequences. This file contains the construction sequence (additions by type and color, and collapses) by each age group and dyad type.
https://doi.org/10.6084/m9.figshare.12159213

Tower heights. This provides the tower heights for each dyad. https://doi.org/10.6084/m9.figshare.12159345

REFERENCES

Amir, D., Jordan, M. R., McAuliffe, K., Valeggia, C. R., Sugiyama, L. S., Bribiescas, R. G., Snodgrass, J. J., & Dunham, Y. (2020). The developmental origins of risk and time preferences across diverse societies. *Journal of Experimental Psychology: General, 149,* 650–661. DOI

Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics, 95*(4), 1438–1443. DOI

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7–15. DOI

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology, 59*(5), 960–970. DOI

Blais, A. R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1,* 33–47. https://ssrn.com/abstract=1301089

Booth, A. L., & Nolen, P. (2012). Gender differences in risk behaviour: Does nurture matter? *The Economic Journal, 122*(558), 56–78. DOI

Boyer T.W., & Byrnes J.P. (2016). Risk-taking. In Levesque R. (Ed.), *Encyclopedia of Adolescence* (pp. 1–5). Cham, Switzerland: Springer. DOI

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review, 101*, 327-341.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin, 125*(3), 367–383. DOI

Caldwell, C. A., & Millen, A. E. (2008). Experimental models for testing hypotheses about cumulative cultural evolution. *Evolution and Human Behavior, 29*(3), 165–171. DOI

Cherney, I. D., & London, K. (2006). Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children. *Sex Roles, 54,* 717–726. DOI

Cobey, K. D., Laan, F., Stulp, G., Buunk, A. P., & Pollet, T. V. (2013). Sex differences in risk taking behavior among Dutch cyclists. *Evolutionary Psychology, 11*(2), 350–364. DOI

Cochard, F., Couprie, H., & Hopfensitz, A. (2016). Do spouses cooperate? An experimental investigation. *Review of Economics of the Household, 14,* 1–26. DOI

Cross, C. P., Cyrenne, D. L. M., & Brown, G. R. (2013). Sex differences in sensation-seeking: A meta-analysis. *Scientific Reports, 3,* 1–5. DOI

Deck, C., Lee, J., Reyes, J., & Rosen, C. (2012). Risk-taking behavior: An experimental analysis of individuals and dyads. *Southern Economic Journal, 79*(2), 277–299. DOI

Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(3),* 709–730. DOI

Fischer, D., & Hills, T. T. (2012). The baby effect and young male syndrome: Social influences on cooperative risk-taking in women and men. *Evolution and Human Behavior, 33*(5), 530–536. DOI

Fischhoff, B., & Kadvany, J. (2011). *Risk: A Very Short Introduction (1st ed).* New York, NY, USA: Oxford University Press. DOI

Fisher, M. L., Meredith, T., & Gray, M. (2018). Sex differences in mental rotation ability are a consequence of procedure and artificiality of stimuli. *Evolutionary Psychological Science, 4*, 124–133. DOI

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution, 7*(11), 1325–1330. DOI

Gardner, M., & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental Psychology, 41*(4), 625–635. DOI

Ginsburg, H. J., & Miller, S. M. (1982). Sex differences in children's risk-taking behavior. *Child Development, 53*(2), 426–428. DOI

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics, 112*(2), 631–645. DOI

Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica, 77*(5), 1637–1664. DOI

Hartup, W. (2009). Friendships in childhood. In H. T. Reis & S. Sprecher (Eds.), *Encyclopedia of Human Relationships* (pp. 722–726). Thousand Oaks, CA, USA: SAGE Publications, Inc. DOI

Jessor, R., & Jessor, S. L. (1977). *Problem Behavior and Psychological Development: A Longitudinal Study of Youth*. New York, NY, USA: Academic Press.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341–350. DOI

Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of Psychology: A Systematic Text in the Science of Behavior* (pp. 357–358). New York, NY, USA: Appleton-Century-Crofts.

Kelley, A. E., Schochet, T., & Landry, C. F. (2004). Risk taking and novelty seeking in adolescence: Introduction to part I. *Annals of the New York Academy of Sciences, 1021*, 27–32. DOI

Kirley M., von der Osten F.B. (2014) Adaptive risk aversion in social dilemmas. In: Dick G. et al. (eds) *Simulated Evolution and Learning. SEAL 2014. Lecture Notes in Computer Science, vol 8886*. Springer, Cham, Switzerland. DOI

Kruger, D. J., & Nesse, R. M. (2004). Sexual selection and the male:female mortality ratio. *Evolutionary Psychology, 2*(1), 66–85. DOI

Leigh, B. C. (1999). Peril, chance, adventure: Concepts of risk, alcohol use and risky behavior in young adults. *Addiction, 94*(3), 371–383. DOI

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*(2), 75–84. DOI

Lüdecke, D. (2018). *sjstats: Statistical functions for regression models. R package version 0.14.1.* Retrieved from https://CRAN.R-project.org/package=sjstats

Lupfer, M., Jones, M., Spaulding, L., & Archer, R. (1971). Risk-taking in cooperative and competitive dyads. *Journal of Conflict Resolution, 15*(3), 385–392. DOI

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex Differences in Antisocial Behaviour: Conduct Disorder, Delinquency, and Violence in the Dunedin Longitudinal Study*. Cambridge, UK: Cambridge University Press.

Nelson, J. A. (2015). Are women really more risk-averse than men? A re-analysis of the literature using expanded methods. *Journal of Economic Surveys, 29*(3), 566–585. DOI

Pawlowski, B., Atwal, R., & Dunbar, R. I. M. (2008). Sex differences in everyday risk-taking behavior in humans. *Evolutionary Psychology, 6*(1), 29–42. DOI

Peters, M., Servos, P., & Day, R. (1990). Marked sex differences on a fine motor skill task disappear when finger size is used as covariate. *Journal of Applied Psychology, 75*(1), 87–90. DOI

Pfefferbaum, B., & Wood, P. B. (1994). Self-report study of impulsive and delinquent behavior in college students. *The Journal of Adolescent Health, 15*(4), 295–302. DOI

R Core Team (2020). *R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing*. Retrieved from https://www.r-project.org/

Rao, N., Esber, A., Turner, A., Chilewani, J., Banda, V., & Norris, A. (2016). The impact of joint partner decision making on obstetric choices and outcomes among Malawian women. *International Journal of Gynecology & Obstetrics, 135*(1), 61–64. DOI

Reindl, E., Apperly, I. A., Beck, S. R., & Tennie, C. (2017). Young children copy cumulative technological design in the absence of action information. *Scientific Reports, 7*, 1–11. DOI

Rosetti, M. F., Valdez, B., & Hudson, R. (2017). Effect of spatial scale on children's performance in a searching task. *Journal of Environmental Psychology, 49*, 86–95. DOI

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703. DOI

Slovic, P. (1966). Risk-taking in children: Age and sex differences. *Child Development, 37*(1), 169–176. DOI

Steinberg, L. (2004). Risk taking in adolescence: what changes, and why?. *Annals of the New York Academy of Sciences, 1021*(1), 51–58. DOI

Stöckl, T., Huber, J., Kirchler, M., & Lindner, F. (2015). Hot hand and gambler's fallacy in teams: Evidence from investment experiments. *Journal of Economic Behavior & Organization, 117*, 327–339. DOI

Subiaul, F., & Stanton, M. A. (2020). Intuitive invention by summative imitation in children and adults. *Cognition, 202*, 104320. DOI

Suijs, J. (2012). *Cooperative Decision-Making Under Risk*. Heidelberg, Germany: Springer Science & Business Media.

Umphlett, N., Brosius, T., Laungani, R., Rousseau, J., Leslie-Pelecky, D. L. (2009). Ecosystem Jenga! *Science Scope, 33*, 57–60

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599–604. DOI

Von Neumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior (2nd rev. ed.)*. Princeton, New Jersey, USA: Princeton University Press.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271–294. DOI

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York, NY, USA: Springer-Verlag.

Wilke, A., Sherman, A., Curdt, B., Mondal, S., Fitzgerald, C., & Kruger, D. J. (2014). An evolutionary domain-specific risk scale. *Evolutionary Behavioral Sciences, 8*(3), 123–141. DOI

Zuckerman, M. (2007). The sensation seeking scale V (SSS-V): Still reliable and valid. *Personality and Individual Differences, 43*(5), 1303–1305. DOI

Zuckerman, M., Kolin, E. A., Price, L., & Zoob, I. (1964). Development of a sensation-seeking scale. *Journal of Consulting Psychology, 28*(6), 477–482. DOI