CREATIVITY
Theories—Research—Applications
Vol. 12, Issue 1, 2025
sciendo

Moral Disengagement as a Main Component of Malevolent Creative Potential

Cecilia Segatta[1], **Tiziana Pozzoli**[2], and **Sergio Agnoli**[1,3]

1 Department of Life Sciences, University of Trieste, Italy, ORCID: 0009-0001-9667-7950
2 Department of Development and Social Psychology, University of Padova, Italy, ORCID: 0000-0001-7474-553X
3 Marconi Institute for Creativity, Italy, ORCID: 0000-0003-3004-7988

ABSTRACT

Malevolent creativity uses creative thinking for harmful or destructive goals, displaying characteristics of originality and effectiveness, but with harmful (i.e., malevolent) intent. In the present work, we hypothesized that one of the main components of creative potential in malevolent scenarios is moral disengagement, which should modulate the association between originality and malevolence in the generation of new ideas. The current study aimed at exploring malevolent creative potential in a sample of 190 participants, both in neutral everyday situations and in scenarios evoking malevolent intentions. The results showed that moral disengagement was indeed associated with a higher rate of malevolent creative behaviors as well as with a higher ability to produce original solutions, especially in scenarios requiring malevolent intentions. Particularly, moral disengagement proved to be the moderator between malevolence and originality of participants' responses. Specifically, in malevolent scenarios, results revealed an association between malevolence and originality but only at medium or high levels of moral disengagement. In conclusion, moral disengagement represents a pivotal resource to succeed in creativity in scenarios eliciting malevolent intentions.

KEYWORDS:

malevolent creativity, moral disengagement, creative potential, divergent thinking, malevolence, originality

Article history:

Received: July 6, 2024
Received in revised form: December 21, 2024
Accepted: January 7, 2025

ISSN 2354-0036
DOI: 10.2478/ctra-2025-0001

Corresponding author at:

Sergio Agnoli
E-MAIL: sergio.agnoli@units.it

INTRODUCTION

The commonly accepted definition of creativity identifies two main requirements for its occurrence: potential originality and effectiveness (Corazza, 2016; Runco & Jaeger, 2012). As suggested by Runco and colleagues (2012), creative ideas need to possess a certain degree of effectiveness; however, without originality (i.e., novelty, authenticity, non-obviousness; see Corazza, 2016; Simonton, 2012), no creativity can occur. Creative potential indeed refers to a latent ability inherent in every individual to generate potentially valuable, effective, and original ideas. This latent ability is conceptualized as the confluence of multiple resources (according to the multivariate approach proposed by Lubart et al., 2013; see also Nelson & Botella, 2017 and Mastria et al., 2018), such as cognitive styles, intelligence, knowledge (i.e., cognitive factors), motivation, personality, and affect (i.e., conative factors), and socio-cultural aspects (i.e., environmental factors). Achievements pursued through exploitation of creative potential have been pivotal in the history of human discoveries in arts, science, and technology (Agnoli et al., 2016; Andreasen & Ramchandran, 2012), contributing to general well-being and to incremental and exponential advancements in human knowledge across different domains. Generally speaking, the progress of the human species presents creativity in a positive light, as a benevolent force. However, when individuals deliberately use their creative potential for less noble intentions, such as for generating original and useful ways to harm others, a dark side of creativity emerges, showing the flip side of the coin: malevolent creativity (Cropley et al., 2008; Baas et al., 2019; Shi et al., 2023).

Malevolent creativity has many different examples in history; famous among them are the cruel but still innovative tortures of the medieval inquisition (Ara, 2015). Several cases in the history of inventions have shown famous inventors generating both positive and negative products: for example, Leonardo Da Vinci created both the most famous painting of the history (The Monna Lisa) as well as the first prototype of a tank. The term "malevolent creativity" has been coined to refer to the use of creative thinking to pursue violent and destructive goals (e.g., acts of terrorism, stealing, lying; Cropley et al., 2008; Cropley et al., 2014). Malevolent creativity is based on the same requirements defining general creativity, namely originality and effectiveness, but with the specific aim of causing harm on a physical or psychological level (Harris & Reiter-Palmon, 2015). As in the case of general creativity (Batey et al., 2009), the potential underlying the expression of malevolent creativity is commonly assessed through divergent thinking tasks (e.g., Alternative Uses Test and Realistic Problems tasks; Lee & Dow, 2011; Gao et al., 2022; Zhao et al., 2022), and scored in terms of fluency, originality, as well as malevolence. Several studies assimilated general creative potential with malevolent creativity potential, arguing that creative individuals in a general domain could be potentially creative in malevolent scenarios as well (Hao et al., 2016; Jonason et al., 2017; Perchtold-Stefan et al., 2020). Other studies, though, highlighted that creative individuals do not necessarily succeed in scenarios requiring malevolent creativity (Keem et al., 2018; Wang, 2019), revealing a complex relationship between creative potential and its translation into malevolent creativity. Recent evidence on the individual differences sustaining malevolent creativity highlighted that in-

dividuals with a propensity for aggressiveness, anger, and antagonism are more likely to generate malevolent creative ideas (Cheng et al., 2021; Harris & Reiter-Palmon, 2015; Perchtold-Stefan et al., 2021, 2022). In the same vein, social aversive personality traits (e.g., Machiavellianism, psychopathy, or Dark Triad traits; Jonason et al., 2017) are pointed to as possible antecedents of malevolent creativity. Moreover, certain social scenarios characterized by social exclusion, provocativeness, and unfairness seem to facilitate malevolent creative responses (Hao et al., 2016; Perchtold-Stefan et al., 2020).

The role of morality in malevolent creative potential and behavior

Beside personal traits and environmental conditions, recent research highlighted how moral aspects, such as moral foundations, can be considered additional factors underlying malevolent creative behaviors (Kapoor & Kaufman, 2022). In fact, in a recent study conducted by Zhao and colleagues (2022), moral reasoning has been identified as a moderating factor between creative potential and malevolent creative behavior (assessed with the Malevolent Creative Behavior Scale by Hao et al., 2016). Results revealed that especially individuals characterized by low moral reasoning and high creative potential were more likely to engage in malevolent creative behaviors (Zhao et al., 2022). Besides moral reasoning, moral disengagement recently emerged to be associated with the expression of malevolent creativity (Shi et al., 2023). Moral disengagement refers to a psychological process through which individuals justify harmful and immoral behaviors towards others (Bandura, 2011). By employing moral justifications, using euphemistic language to downplay harmful actions, favorably comparing one's actions to those of others, shifting personal responsibility onto others, diffusing responsibility within a group, minimizing or distorting the consequences of one's actions, dehumanizing others to justify mistreatment, and blaming victims, individuals can avoid experiencing negative self-sanctions. This selective use of moral disengagement mechanisms can increase the propensity to engage in unethical behaviors without experiencing remorse, guilt, or shame (Bandura, 2016). According to the social cognitive theory (Bandura, 1989), moral disengagement's functioning is characterized by a reciprocal determinism of personal, environmental, and behavioral aspects. Shi and colleagues (2023) provided evidence of how the influence of malevolent creative performance on malevolent creative behavior is moderately mediated by moral disengagement. In particular, the authors highlighted how moral disengagement mediates the relation between the ability to generate malevolent creative ideas and the engagement in malevolent creative behaviors in real life. However, delving into this mediation, a secondary moderation effect of the individual's conscience over the relation between moral disengagement and malevolent creative behavior scale has been identified (Shi et al., 2023; Tillman et al., 2018). In other terms, individuals with a strong sense of responsibility were less inclined to transpose their malevolent creative ability into malevolent creative behaviors, even at high levels of moral disengagement. On the contrary, individuals with a weaker sense of responsibility were more likely to engage in malicious behaviors, decentralizing and transferring responsibility for their actions (Kapoor & Kaufman, 2022).

Research questions and hypotheses

To summarize, previous studies depict malevolent creativity as a result of the interaction of personality, situational, and emotional aspects, with morality appearing to play a fundamental role. In fact, moral disengagement, as much as moral reasoning and conscience were pivotal aspects of malevolent creative expressions. However, it remains unclear whether the ability to disengage from moral rules is a necessary condition for the expression of creative potential in non-benevolent scenarios (i.e., malevolent situations). Starting from previous literature, we further question: can moral disengagement predict success in creative tasks, especially in situations requiring malevolence? Whereas the previous literature (Shi et al., 2023) was interested in understanding the role of moral disengagement in the relationship between the ability to generate malevolent creative ideas and the engagement in malevolent creative behaviors in real life, with the present study we instead want to understand whether moral disengagement is a fundamental mechanism for the ability to generate malevolent creative ideas (i.e., malevolent creative potential). We therefore tried to understand whether creative potential could vary depending on the context and specifically whether creative potential includes morality (or the lack of it) as a necessary component, in particular in malevolent scenarios. In other terms, do individuals with high creative potential in everyday, general scenarios achieve equal success in creativity also when harmful intents are pursued? Or can moral disengagement be instead considered a key mechanism distinguishing general creative potential from the potential expressed in malevolent scenarios? Considering that succeeding in malevolent creativity requires producing ideas that are both original and malevolent, we hypothesize that moral disengagement could act as a sort of bond between the ability to produce original ideas and the ability to produce malevolent ideas.

In the present study we explored the ability to produce ideas in everyday situations and in social-unfair situations (i.e., malevolent situations). We hypothesize (H1) that the tendency to exhibit malevolent creativity behaviors in real-life would predict creative performance, especially in malevolent scenarios, as highlighted by previous research (Shi et al., 2023; Zhao et al., 2022). Moreover, we hypothesized (H2), confirming results of past research (Shi et al., 2023), that moral disengagement could serve as a crucial mechanism for succeeding in creative tasks (i.e., as a key predictor of the ability to produce original response), especially in situations requiring harmful intentions; in other terms, moral disengagement might be intended as a main component of creative potential in malevolent scenarios. More specifically, we hypothesized (H3) that moral disengagement could predict creative performance in malevolent scenarios by moderating the association between the ability to produce original responses and the ability to produce malevolent responses in these situations. While high levels of moral disengagement could allow individuals to produce both original and malevolent ideas, low levels of moral disengagement could instead hamper the association between originality and malevolence, thus preventing individuals from producing ideas that are both original and malevolent.

METHOD

Participants

Two hundred and twenty-four participants took part in an online study. The data of a total of 34 participants were excluded for different reasons: nine participants have been excluded due to their execution times, which were lower than 17 minutes (i.e., lower than the minimum time required for performing all tasks); 7 participants have been excluded due to their execution times, which were higher than 1 hour (i.e., more than the 4*SD* from the average execution time). Additionally, 18 participants did not complete the entire protocol. The final sample consisted in 190 adults (155 females; $M_{age} = 22.8$, $SD_{age} = 6.3$) recruited at the University of Trieste who received course credits for their participation. Sample size was computed starting from the effect between moral disengagement and creative potential ($R^2 = 0.14$) that emerged in the paper by Shi et al. (2023). Using the software G*Power 3.1.9.7, taking into account the number of predictors in our models, we estimated that a minimal sample size of 110 participants would be sufficient to observe an effect of $f^2 = 0.163$ with a .95 power (critical $F = 2.69$).

To participate in the study, each individual had to complete informed consent documentation. Participants were involved in an online survey, containing questionnaires and divergent-thinking problems. All data were collected and processed ensuring anonymity and participants were free to discontinue the study at any moment without giving any explanation. The study conformed to the Declaration of Helsinki and was approved by the Ethical Committee of the University of Padova.

Instruments

Malevolent Creativity Task (MCT). The Malevolent Creativity Task is a set of open-ended realistic problems aimed at eliciting malevolent use of creativity (Hao et al., 2016; Perchtold-Stefan et al., 2020). The task is based on a series of scenarios that describe different sorts of unfair behaviors involving peers and associates producing a damage or a disadvantage to the main character, the participant. Participants are instructed to generate as many original ideas as possible to respond to example unfair situations, with the aim to revenge against the peers or the associates. In the present study the following two problems were selected:

> **Task 1**: "*Your neighbor asks you to help him with renovations in his flat and offers to pay you for your troubles. Since you are currently low on money, you agree. After the work is done, you ask him for the payment he promised. However, your neighbor insists that such an agreement never took place and you just imagined the whole thing*." (Perchtold-Stefan et al., 2020);
>
> **Task 2**: "*You're going downtown when you see a friend of yours jogging quickly in your direction. It's clear that your friend didn't see you, but before you can do anything, you collide, and you both fall to the ground. You land heavily on your backpack and hear something shatter. Your friend seems very angry with you. He gets up quickly, doesn't apologize, doesn't offer to help you, and claims it's all your fault. As soon as you get up and look in your backpack, you realize that your laptop is ruined. You want to get back at that friend for bumping into you, being rude, and especially for breaking your computer, but you don't want to get caught*." (Hao et al., 2016).

Realistic Problems Task (RPT). Two open-ended problems were selected from the Realistic Problems Task (Agnoli et al., 2016; Runco et al., 2005), instructing participants to generate as many different solutions as possible. The task is based on open-ended problems describing neutral situations that participants may actually experience in their daily life. Again, participants were asked to produce as many original solutions as possible for each of the following problems:

> **Task 3**: "*You want to consult the meaning of a word you don't know in the dictionary. Once you spot the dictionary on the top shelf of the bookcase, out of reach, you remember that your ladder is broken. What objects can you use or in what different ways can you try to get the dictionary, without the ladder?*";
>
> **Task 4**: "*The weekend is here and you finally have time to go out with your dog for a long walk, but 30 minutes from home you remember 3-4 things you need to buy at the supermarket before dinner tonight. You forgot these things once before and you promised yourself you wouldn't forget them again. You have neither a pen nor a pencil with you. What can you do?*".

Divergent thinking scoring. All responses to the problems of both tasks (MCT and RPT) were scored for originality, malevolence, and fluency (i.e., number of responses produced by a participant). Originality has been scored by two independent trained raters judging the responses on the basis of the criteria reported in Silvia and colleagues (2008). Specifically, responses of each problem were first listed alphabetically and participants identification codes were hidden from the raters. This procedure ensured that raters were blind to the serial position of the responses, to the total number of responses, and to the responses order. Raters read all the responses to the given problem before scoring them and they scored responses separately. Responses received a rating form 1 slated for "not original at all" to 5 "very original," considering that an original response should be uncommon, remote, and clever. Raters were instructed to consider all three criteria in their ratings and in particular to consider that a strength in one dimension could offset weaknesses in another dimension (Silvia et al., 2008). Finally, the ratings of the two judges were averaged for each solution. The interclass correlation coefficient (ICC) between the two judges in the originality ratings was very good: ICC = 0.98.

Malevolence was scored using an adapted version of the Positivity versus Negativity Solution Evaluation Scale (PNSES; Harris et al., 2013). This scoring allowed to separate the judgement on originality from the judgement on the negativity (i.e., harmfulness) of a solution. Negativity was indeed associated with the harmful nature of the response. On the contrary, a positive solution to the problem avoided to use physical or mental harm against persons or objects. In the original version of the PNSES, solutions were judged in three steps: 1) a categorical (1 = positive; 2 = negative) judgement of positivity *vs.* negativity; 2) a 1–6 rating (from 1 = highly positive to 6 = highly negative) of the degree of positivity/negativity; as well as 3) a consensual revision of the scores. In order to conform the PNSES scale to the originality scale used in the scoring of originality, in the present study we adapted the second step using a 1–5 Likert scale (from 1 = highly positive to 5 = highly negative). Again, two independent, trained judges rated each solution. ICC was very good: 0.94. Finally, the ratings of the two judges were averaged for each solution.

Malevolent Creativity Behavior. In order to measure individual attitude to perform malevolent behaviors in daily life, we used the Malevolent Creativity Behavior Scale (MCBS– Hao et al., 2016). The scale is a 13-item self-report questionnaire, which investigates an individual's tendency to ideate MC actions on three dimensions: hurting people, lying, and playing tricks. Participants were instructed to provide answers to 13 statements (e.g., *how often do you play tricks on people as revenge?*) on a 5-point Likert scale (from 0 = never, to 4 = usually), indicating the frequency of each behavior in their daily life. The reliability of the scale was good: $\alpha = 0.90$.

Moral Disengagement. Moral disengagement was measured through the Moral Civil Disengagement Scale (MCD– Caprara et al., 2006). Participants were asked to rate the extent to which they agreed (1 = Strongly disagree to 5 = Strongly agree) with each of the 32 items (e.g., "*It's right to come to blows to protect your friends*"). The scale encompassed all the moral disengagement mechanisms outlined by Bandura (2016). Participants' responses were averaged to form a global score ($\alpha = 0.82$); the higher the score, the higher the proneness to moral disengagement.

PROCEDURE

The study was implemented online using the PsyToolkit platform (Stoet, 2010, 2017). Participants started the study by answering two divergent thinking problems, (i.e., a neutral problem and a malevolent problem). The presentation order of the two types of problem was randomized. Participants were instructed to produce as many original solutions as possible for each of the problems in 3 minutes. Before presenting the problem, an example of an open-ended problem was presented to the participants with a list of possible original solutions. After the two tasks, participants were asked to answer the MCD questionnaire (Caprara et al., 2006). Following that, other two divergent thinking problems were presented (again, one neutral and one malevolent problem, in a randomized order) and finally the MCBS scale (Hao et al., 2016) was presented, followed by demographic questions (i.e., age and gender). In total, the study lasted around 25 minutes.

RESULTS

Preliminary analyses

As a first preliminary analysis, to ensure plausibility of the data and especially to identify careless participants and responses, we computed the Intra-individual Response Variability (IRV) for the responses provided by our participants in the two questionnaires used in the study (MCBS and MCD) using the R *careless* package by Yentes and Wilhelm (2018). IRV is the "standard deviation of responses across a set of consecutive item responses for an individual" (Dunn, Heggestad, Shanock, & Theilgard, 2018, p. 108). It has been identified as a reliable indicator of insufficient effort respond-

ing and of degraded response quality. Specifically, as suggested by Dunn et al. (2018), we used IRV to identify participants with low IRV values as outliers, which reflect straightlining responses. This effect usually emerged in the final items of the responses in careless responders. We computed both a total IRV for each participant in MCBS and in MCD and IRV values for four consecutive subsets of items in the two questionnaires (this method allows to identify whether careless responses have been provided in different parts of the questionnaires, especially in the last part). No participant emerged as outlier from these analyses, ensuring a good response quality and plausibility of the data.

Correlations among the study variables were then explored (Table 1). The analysis showed that moral disengagement was positively correlated with self-reported malevolent creativity (as measured through MCBS; $r = .52$, $p < .001$). In addition, moral disengagement resulted as positively correlated with the originality of solutions both in the Malevolent Creative Task ($r = .42$, $p < .001$) and, even if at a lower degree, in the Realistic Problem Task ($r = .272$, $p < .001$). Self-reported malevolent creativity behavior positively correlated with originality in Malevolent Creative Task ($r = .332$, $p < .001$). Last, originality in Realistic Problem Task was strongly correlated with originality in Malevolent Creative task ($r = .672$, $p < .001$).

Table 1. Correlations and descriptives (means and standard deviations) of the study variables

	MD	SMC	Orig. RPT	Orig. MCT	Malev. RPT	Malev. MCT
MD	-					
SMC	0.520***	-				
Orig. RPT	0.272***	0.153*	-			
Orig. MCT	0.420***	0.332***	0.672***	-		
Malev. RPT	0.201**	0.152*	0.354***	0.429***	-	
Malev. MCT	0.328***	0.211**	0.402***	0.559***	0.207**	-
M	1.60	24.1	1.13	1.21	3.02	3.81
SD	0.322	8.62	0.246	0.472	0.0840	0.312
Min	1.06	13.0	1.00	1.00	2.79	2.75
Max	2.91	62.0	2.60	5.00	3.47	5.00

Notes: * $< p$ .05, ** $< p$ .01, *** $< p$ .001; MD = Moral Disengagement; Orig. = Originality; Malev. = Malevolence; SMC = Self-reported Malevolent Creativity; RPT = Realistic problem Task; MCT = Malevolent Creativity Task

Predictors of malevolent creativity

The link of malevolent creativity behavior with response originality in neutral and malevolent scenarios

As a first analysis, to test H1, we explored the association between the tendency to perform malevolent creative behaviors and performance in the production of original solutions in the two divergent thinking tasks (neutral and malevolent). A Generalized Linear Mixed Models (GLMM)

was tested using the SPSS v. 26 software: originality was tested as a dependent variable and MCBS scores, problem type, and the interaction between MCBS and problem type were used as fixed independent variables, while subjects and the interaction subjects*problems were introduced in the model as random effects. To control for possible effects of outliers and violations of normality assumptions, we used a robust error estimation.

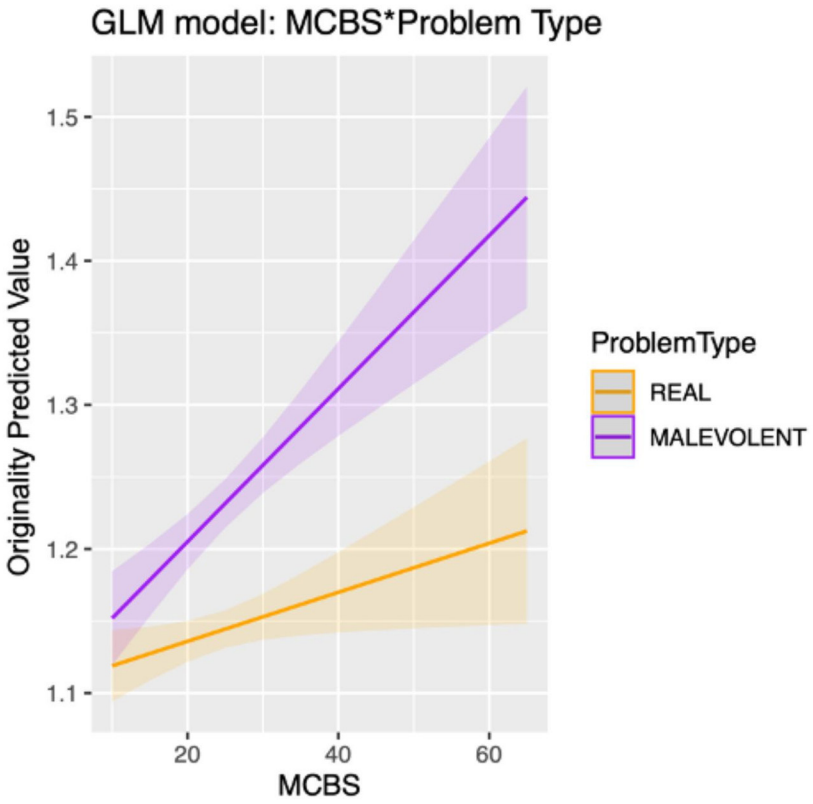

Figure 1. The predictive power of the attitude to perform malevolent creative behaviors (i.e., Malevolent Creative Behavior Scale) on the originality of solutions produced in the two divergent thinking tasks (realistic-neutral and malevolent).

The model showed that the interaction between MCBS and Problem Type approached significance in predicting originality, $F(1, 3037) = 3.804$, $p = .051$ (Table 2a). No main effect of MCBS, $F(1,3037) = 2.880$, $p = .090$, and Problem Type, $F(1, 3037) = 1.341$, $p = .247$, was found. In order to unravel the interactive effect between malevolent creative behavior and problem type, we tested two separate models exploring separately the effect of malevolent creative behavior on response originality in the two problem scenarios (neutral and malevolent; Table 2b); results revealed that MCBS scores tend to significantly predict the ability to produce original solutions only in the malevolent context, but not in neutral scenarios (Figure 1).

Table 2a. Effects of the linear mixed model, testing the predictive role of creative malevolent behaviors and of the type of problem (neutral vs. malevolent) on participants' response originality.

						95% CI	
Random Effects	**b**	**SE**	**Z**			**Lower**	**Upper**
Intercept subject	.068	.009	7.228***			.214	.240
Intercept subject*problem type	.029	.005	5.635***			.021	.042
Residual	.226	.007	34.565***			.214	.240
						95% CI	
Fixed Coefficients	**b**	**SE**	**t**	**df1**	**df2**	**Lower**	**Upper**
Intercept	.898	.1598	5.617***	3	3037	.584	1.211
Problem Type (RPT vs MCT)	.098	.0844	1.158	1	3037	-.068	.263
MCBS	.013	.0069	1.845	1	3037	-.001	.026
MCBS* Problem Type (RPT vs MCT)	-.007	.0036	-1.950+	1	3037	-.014	.000

+ $p = .051$; *$p < .05$, ** $p < .01$, ***$p < .001$

Note: SE = Standard Error; df = degrees of freedom; CI = Confidence Interval; MCBS = Malevolent Creative Behavior Scale; RPT = Realistic Problem Task; MCT = Malevolent Creativity Task

Table 2b. Effects of the linear mixed models, testing the effect of creative malevolent behaviors on the response originality in the neutral (1) and in the malevolent scenarios (2).

1. Neutral scenarios

						95% CI	
Random Effects	**b**	**SE**	**Z**			**Lower**	**Upper**
Intercept subject	.041	.007	6.280***			.030	.056
Intercept subject*problem type	.001	.004	.232			.000	3.896
Residual	.210	.007	28.586***			.196	.225
						95% CI	
Fixed Coefficients	b	SE	t	df1	df2	Lower	Upper
Intercept	1.047	.0811	12.912***	1	1915	.888	1.206
MCBS	.003	.0034	1.014	1	1915	-.003	.010

2. Malevolent scenarios

						95% CI	
Random Effects	**b**	**SE**	**Z**			**Lower**	**Upper**
Intercept subject	.110	.022	5.098***			.075	.162
Intercept subject*problem type	.071	.016	4.462***			.046	.111
Residual	.252	.012	20.231***			.228	.277
						95% CI	
Fixed Coefficients	**b**	**SE**	**t**	**df1**	**df2**	**Lower**	**Upper**
Intercept	.862	.1723	5.004***	1	1122	.524	1.200
MCBS	.014	.0075	1.903+	1	1122	.000	.029

+ $p = .057$; *$p < .05$, ** $p < .01$, ***$p < .001$

Note: SE = Standard Error; df = degrees of freedom; CI = Confidence Interval; MCBS = Malevolent Creative Behavior Scale

The role of moral disengagement on response originality in different scenarios

In a further GLM model, to test H2, we explored the predictive power of moral disengagement on the originality of solutions produced in the two divergent tasks. Again, originality was introduced as a dependent variable in the model, while MCD scores, problem type (i.e., neutral *vs.* malevolent) and the interaction between MCD scores and problem type were tested as independent fixed variables. Again, we tested the random effect of subjects and of the subject*problems interaction as well as used a robust estimation of the error.

This model highlighted the main effect of moral disengagement in predicting originality, the main effect of Problem Type, and an effect of the interaction between moral disengagement and Problem Type, $F(1, 3037) = 5.778$, $p = .016$ (Table 3a). In order to unravel the interactive effect between moral disengagement and problem type, we tested two separate models exploring separately the effect of moral disengagement on response originality in the two problem scenarios (neutral and malevolent; Table 3b); results revealed that moral disengagement predicted significantly the ability to produce original solutions in both scenarios, even if its predictive power in malevolent scenarios remains significatively higher in comparison to neutral scenarios as testified by the interactive effect in the first main model (Table 3a) and as shown in Figure 2.

Table 3a. Effects of the linear mixed model, testing the predictive role of moral disengagement and of the type of problem (neutral *vs.* malevolent) on participants' response originality

						95% CI	
Random Effects	**b**	**SE**	**Z**			**Lower**	**Upper**
Intercept subject	.061	.009	7.056***			.046	.080
Intercept subject*problem type	.028	.005	5.525***			.020	.041
Residual	.226	.007	34.565***			.214	.240
						95% CI	
Fixed Coefficients	**b**	**SE**	**t**	**df1**	**df2**	**Lower**	**Upper**
Intercept	.365	.3005	1.216	3	3037	-.224	.955
Problem Type (RPT vs MCT)	.395	.1853	2.132*	1	3037	.032	.758
MCD	.524	.1980	2.646**	1	3037	.136	.912
MCD* Problem Type (RPT vs MCT)	-.292	.1214	-2.404*	1	3037	-.530	-.054

*$p < .05$, ** $p < .01$, ***$p < .001$

Note: SE = Standard Error; df = degrees of freedom; CI = Confidence Interval; MCD = Moral Civil Disengagement Scale; RPT = Realistic Problem Task; MCT = Malevolent Creativity Task

Table 3b. Effects of the linear mixed models, testing the effect of moral disengagement on the response originality in neutral (1) and malevolent scenarios (2)

1. Neutral scenarios

						95% CI	
Random Effects	**b**	**SE**	**Z**			**Lower**	**Upper**
Intercept subject	.039	.006	6.166***			.028	.053
Intercept subject*problem type	.001	.004	.284			.000	1.017
Residual	.209	.007	28.617***			.196	.224
						95% CI	
Fixed Coefficients	b	SE	t	df1	df2	Lower	Upper
Intercept	.826	.1330	6.211***	1	1915	.565	1.087
MCD	.190	.0873	2.177*	1	1915	.019	.361

2. Malevolent scenarios

						95% CI	
Random Effects	**b**	**SE**	**Z**			**Lower**	**Upper**
Intercept subject	.093	.020	4.646***			.061	.142
Intercept subject*problem type	.073	.016	4.506***			.047	.112
Residual	.251	.012	20.275***			.228	.277
						95% CI	
Fixed Coefficients	**b**	**SE**	**t**	**df1**	**df2**	**Lower**	**Upper**
Intercept	.325	.3059	1.063	1	1122	-.275	.925
MCD	.551	.2020	2.730**	1	1122	.155	.948

*$p < .05$, ** $p < .01$, ***$p < .001$

Note: SE = Standard Error; df = degrees of freedom; CI = Confidence Interval; MCD = Moral Civil Disengagement Scale

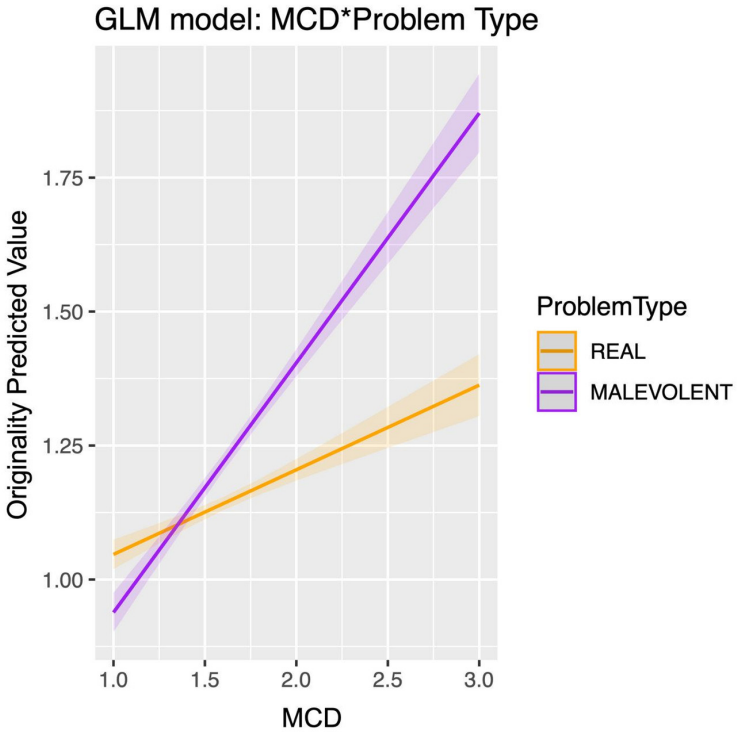

Figure 2. The predictive power of moral disengagement on the originality of solutions produced in the two divergent thinking tasks (realistic-neutral and malevolent)

Moral disengagement as moderator between originality and malevolence

To test whether moral disengagement could moderate the relationship between the ability to produce original responses and the malevolence of the responses (H3), a further GLMM was performed. In this model, originality was tested as a dependent variable, while moral disengagement and response malevolence (as measured through the PNSES scale) were used as fixed factors. Moreover, we tested the interaction between malevolence and moral disengagement and the interaction between malevolence, moral disengagement, and type of problems (neutral *vs*. malevolent). Again, subjects and subjects*problems were tested as random effects, and we used a robust error estimation.

The model showed a main effect of moral disengagement. Concurrently, malevolence did not emerge as significant. These effects were further specified by an interaction between malevolence and moral disengagement and more specifically by the interaction between malevolence, moral disengagement, and problem type (Table 4a).

To unravel this complex three-way interaction, the interactive effects between malevolence and moral disengagement were explored in two separated GLM models, one for the realistic-neutral situations and the other one for the malevolent situations. The first GLM model investigating the relation between malevolence and originality in neutral problems did not highlight any significant main or interaction effect [$Fs < 1.253$, $ps > .263$].

The second GLM model investigating the relation between response malevolence and response originality in malevolent scenarios highlighted the main effect of malevolence, the main effect of moral disengagement, and an effect of the interaction malevolence and moral disengagement, $F(1, 1120) = 6.506$, $p = .011$ (Table 4b). Specifically, a simple slopes analysis revealed that higher levels of malevolence in the responses predicted higher levels of originality only when associated with a medium (*average*) or a high level of moral disengagement (*+1 SD*), but not at a low moral disengagement level (*-1 SD*; see Table 4c, Figure 3). In scenarios requiring malevolence, moral disengagement acted therefore as a moderator variable between the ability to produce malevolent responses and the ability to produce original responses.

Table 4a. Effects of the linear mixed model, testing the predictive role of response malevolence, moral disengagement, and type of problem (neutral *vs.* malevolent) on participants' response originality

						95% CI	
Random Effects	**b**	**SE**	**Z**			**Lower**	**Upper**
Intercept subject	.044	.007	6.581***			.033	.060
Intercept subject*problem type	.022	.005	4.852***			.015	.033
Residual	.223	.006	34.822***			.211	.236
						95% CI	
Fixed Coefficients	**b**	**SE**	**t**	**df1**	**df2**	**Lower**	**Upper**
Intercept	2.433	.7848	3.101**	4	3036	.895	3.972
Malevolence	-.516	.2711	-1.904	1	3036	-1.048	0.15
MCD	-1.338	.4873	-2.745**	1	3036	-2.293	-.382
Malevolence*MCD	.471	.1690	2.784**	1	3036	.139	.802
Malevolence*MCD* Problem Type (RPT *vs* MCT)	.025	.0095	2.608**	1	3036	.006	.043

*$p < .05$, ** $p < .01$, ***$p < .001$

Note: SE = Standard Error; df = degrees of freedom; CI = Confidence Interval; MCD = Moral Civil Disengagement Scale; RPT = Realistic Problem Task; MCT = Malevolent Creative Task

Table 4b. Effects of the linear mixed model, testing the predictive role of response malevolence and moral disengagement on participants' response originality in malevolent scenarios

			Malevolent scenarios				
						95% CI	
Random Effects	**b**	**SE**	**Z**			**Lower**	**Upper**
Intercept subject	.036	.015	2.435*			.016	.080
Residual	.240	.012	20.341***			.218	.264
						95% CI	
Fixed Coefficients	**b**	**SE**	**t**	**df1**	**df2**	**Lower**	**Upper**
Intercept	4.357	1.7646	2.469*	3	1120	.898	7.820
Malevolence	-.987	.4910	-2.010*	1	1120	-1.950	-.023
MCD	-2.763	1.1656	-2.370*	1	1120	-5.050	-.476
Malevolence*MCD	.821	.3221	2.551*	1	1120	.190	1.453

*$p < .05$, ** $p < .01$, ***$p < .001$

Note: SE = Standard Error; df = degrees of freedom; CI = Confidence Interval; MCD = Moral Civil Disengagement Scale

Table 4c. Simple slope analysis showing the effect of Malevolence on Originality at *-1SD*, average, and *+1SD* values of Moral Disengagement as specified by the scores in the Moral Civil Disengagement Scale (MCD)

					95% CI	
	MCD	**b**	**SE**	**t**	**Lower**	**Upper**
- 1 SD	-.3259	.1615	.0911	1.7732	-.0182	.34
Average	.0000	.6655	.0745	8.9275***	.5184	.81
+ 1 SD	.3259	1.1694	.0860	13.5914***	.9996	1.33

*$p < .05$, ** $p < .01$, ***$p < .001$

Note: SE = Standard Error; CI = Confidence Interval; MCD values are mean-centered

Figure 3. The moderation effect of moral disengagement in the relationship between malevolence and originality of the ideas produced in malevolent scenarios

DISCUSSION

Morality is a relatively new hotspot of interest in malevolent creativity research. Past research focused on the role of a person's moral code over not only the ability to think and engage in malevolent actions, such as lying, but also on the perception of malevolent creativity itself (Cropley et al., 2014). Recently, the moral structure of an individual, in terms of conscientiousness, moral reasoning, and moral disengagement has been investigated, highlighting how lower moral foundations or more flexible morality lead to a greater tendency to engage in malevolent creative attitudes (see Kapoor & Kaufmann, 2022; Shi et al., 2023; Zhao et al., 2022). The present study focused specifically on the malevolent creative potential, exploring the predictive role of the attitude to engage in malevolent creative behaviors and of moral disengagement in its expression. Morality plays a central role in this study, in which we hypothesized that moral disengagement is a main determinant of malevolent creativity potential and that it acts as a moderator between the ability to generate malevolent responses and the ability to generate original responses. Our results confirm this hypothesis, highlighting how the tendency to justify harmful and immoral behaviors towards others (i.e., moral disengagement) is the main mechanism regulating the relationship between the production of malevolent responses and the production of original responses, acting as a bond

between the two necessary requirements (i.e., malevolence and originality) defining a malevolent creative response.

The predictive power of malevolent creative behavior on creative performance.

The first question investigated in this study is whether the attitude to engage in malevolent creative behaviors predicts originality of the solutions in the two divergent thinking tasks (malevolent and realistic-neutral problems). The results seem to confirm our first hypothesis (H1), highlighting a trend in the data that showed how individuals with the tendency to engage in malevolent creative behaviors (i.e., MCBS) were also advantaged in the production of creative (i.e., original) solutions, especially in malevolent scenarios. As mentioned before, this finding is consistent with past research, in which self-reported malevolent creative behavior in daily life resulted positively correlated with malevolent creativity performance on the MCT (Malevolent Creative Task, Perchtold-Stefan et al., 2021).

Moral disengagement as a crucial mechanism for succeeding in malevolent creativity tasks

Our second hypothesis aimed at exploring the role of moral disengagement in the creative performance in divergent thinking tasks set in neutral scenarios and in malevolent scenarios. In line with previous evidence (see Shi et al., 2023), our results show that highly morally disengaged individuals (i.e., MCD) engage more likely in malevolent creative behaviors (i.e., MCBS) and have better performance in terms of originality in the divergent thinking tasks. On the contrary, individuals characterized by low moral disengagement engage less in malevolent creative behaviors and have emerged to exhibit poorer performance in the divergent thinking tasks. Especially, this distinction was clearly evident in malevolent scenarios in which moral aspects are pivotal. Moral disengagement indeed predicted a higher ability to produce original solutions in malevolent scenarios in comparison to neutral scenarios, confirming our second hypothesis (H2). Hence, we found confirmation of the results, which emerged in a recent study by Zhao and colleagues (2022), who highlighted a stronger effect of morality over creativity (i.e., originality and fluency scores) in malevolent scenarios than in neural ones. Moreover, the authors proved that individuals characterized by high moral reasoning tend to produce less original outcomes in specific scenarios of moral injustice, which demand a certain degree of malevolence. Thus, our results confirmed the importance of morality in malevolent creative potential, highlighting that high levels of moral disengagement lead to a better creative performance, especially of a malevolent type.

Moral disengagement as a moderator between originality and malevolence

Furthermore, and more importantly, in the current study we aimed to provide a possible explanation of the role of moral disengagement in malevolent creativity, by exploring its role in the ability to produce ideas that are both original and malevolent. A complementary level of originality and malevolence is indeed crucial to achieving high performance in malevolent creative tasks. The re-

sults of the current study showed that in scenarios encouraging malicious behaviors, high levels of moral disengagement were likely to lead to a greater capacity to produce ideas that are malevolent and original at the same time. On the contrary, low levels of moral disengagement emerged to hamper the ability to produce responses that were both malevolent and original, leading to a failure in creative tasks requiring malevolence (i.e., MCT). At low level of moral disengagement, in particular, malevolent responses were not associated with originality in the production of solutions to malevolence-inducing situations (i.e., MCT), whereas medium to high levels of moral disengagement were necessary to associate malevolence with originality in the production of alternative solutions to the same situations.

These results are in line with past research indicating that individuals with a greater inclination to morally disengage more likely engage in immoral actions across a wide range of situations (Moore, 2015). This aligns with the positive association between moral disengagement and the greater tendency to produce malevolent solutions (observed in this study), in both neutral and malevolent scenarios. It is possible to hypothesize that when a task explicitly requires generating negative and vengeful actions, individuals with high moral disengagement, who comply with the request and produce more malevolent responses, will also be more likely to exhibit greater originality. This may be due to their greater propensity to act immorally and conceive immoral actions without experiencing negative feelings, which can be translated from real life and used as inspiration to tackle the task: having a higher expertise in malevolent actions could help them to understand which kind of malevolent alternative could be more original. This could also explain the higher originality shown by these participants compared to those who, due to greater difficulty in morally disengaging and consequently a lower tendency to commit immoral actions in their daily life, will have more difficulty being original even when trying, because it is required by the situation, to produce malevolent solutions.

In individuals with high moral constraints, the dimensions of malevolence and originality emerged to be unrelated, thus proving the hampering role of morality in establishing an association between malevolence and originality of the responses (confirming once again our H3). Without moral disengagement, individuals seem not to be able to produce ideas that are both high in originality and in malevolence, and therefore they could not potentially culminate in malevolent creativity. Moreover, it is worth highlighting that whereas moral disengagement emerged to be an essential variable to succeeding in potentially harmful situations, its role was essentially irrelevant in neutral situations.

The findings that emerged in the present study may also inform education and intervention practices on which mechanisms we should train in order to enhance the social benefit of creative potential: educative interventions aimed at increasing adherence to moral and ethical social constraints could indeed potentially drive channelization of creative potential toward more positive outcomes.

Limitations and further investigations

It is however worth to highlight some limitations of the present study. First, this research is based on written scenarios describing situations in neutral and in unfair/malevolent scenarios. The ideas produced by participants could not correspond to the behavior they would express in real situations. In addition, we could not exclude the effect of social desirability in participants' responses, which could have reduced the level of intentional malevolence in their responses. Moreover, even if we checked for careless responders, we cannot totally exclude inattentive responses in the self-reported measures, since we did not use specific attention checks in the online questionnaires. It is also worth highlighting that despite the results of past research showing the influence of personality dispositions (e.g., Machiavellianism, psychopathy, or Dark Triad traits; Jonason et al., 2017) in predicting malevolent creativity, in this work we explored only a limited set of possible predictors of this behavior. We therefore call for future studies integrating the effects that emerged in the present study into more complex predictive models, also including personality dispositions and gender differences (which we cannot explore in the present study, because of the gender-unbalanced sample). Lastly, a measurement of general creative behavior (e.g., Runco's Ideational Behavior Scale; Runco et al., 2001) was missing in our protocol; introducing this element in a future study could indeed lead to interesting considerations on eventual differences in the relationship between morality and malevolent vs. neutral creative behaviors.

Despite these limitations, the current study highlighted how moral disengagement plays a key role as a moderating mechanism between the ability to produce malevolent responses and the ability to produce original responses, in vicious scenarios eliciting malevolence. In conclusion, moral disengagement represents a pivotal resource to succeeding in creativity in scenarios eliciting malevolent intentions, thus acting as a main resource for the malevolent creative potential.

Data availability

The complete database is available at the following link:
(https://osf.io/3ch7s/?view_only=ff9fe2cd923f4568946452ea23475862).

Acknowledgements

We would like to thank Michela Bossi for her help in the data collection for this study.

References

Agnoli, S. (2024). IDHOL, a holistic model for the analysis of individual differences in the creative process. *Creativity Research Journal*, 1–11.

Agnoli, S., Corazza, G.E., & Runco, M.A. (2016). Estimating creativity with a multiple-measurement approach within scientific and artistic domains. *Creativity Research Journal*, *28*(2), 171–176.

Andreasen, N.C. & Ramchandran, K. (2012). Creativity in art and science: are there two cultures? *Dialogues in clinical neuroscience*, *14*(1), 49–54.

Ara, R. (2015). *Unruly Practices: The History of Torture throughout the Ages*. Dhaka Univ. Stud. Part F, 26, 75.

Baas, M., Roskes, M., Koch, S., Cheng, Y., & De Dreu, C.K. (2019). Why social threat motivates malevolent creativity. *Personality and Social Psychology Bulletin*, *45*(11), 1590–1602.

Bandura, A. (2011). Moral disengagement. In D.J. Christie (Ed.), *The encyclopedia of peace psychology*. Blackwell Publishing Ltd.

Bandura, A. (2016). *Moral Disengagement: How People Do Harm and Live with Themselves*. Macmillan.

Bandura, A. (1989). Social cognitive theory. In R. Vasta (Ed.), *Annals of child development. Vol. 6. Six theories of child development* (pp. 1–60). JAI Press.

Batey, M., Chamorro-Premuzic, T., & Furnham, A. (2009). Intelligence and personality as predictors of divergent thinking: The role of general, fluid and crystallised intelligence. *Thinking Skills and Creativity*, *4*(1), 60–69.

Caprara, G.V., Barbaranelli, C., Pastorelli, C., Iafrate, C., Beretta, M., Steca, P., & Bandura, A. (2006). La misura del disimpegno morale nel contesto delle trasgressioni dell'agire quotidiano. *Giornale Italiano di Psicologia*, *33*(1), 83–106.

Cheng, R., Lu, K., & Hao, N. (2021). The effect of anger on malevolent creativity and strategies for its emotion regulation. *Acta Psychologica Sinica*, *53*(8), 847.

Corazza, G.E. (2016). Potential originality and effectiveness: The dynamic definition of creativity. *Creative Research Journal*, 28 (3), 258–267

Cropley, D.H., Kaufman, J.C., & Cropley, A.J. (2008). Malevolent creativity: A functional model of creativity in terrorism and crime. *Creativity Research Journal*, *20*(2), 105–115.

Cropley, D.H., Kaufman, J.C., White, A.E., & Chiera, B.A. (2014). Layperson perceptions of malevolent creativity: The good, the bad, and the ambiguous. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(4), 400–412.

Dunn, A.M., Heggestad, E.D., Shanock, L.R., & Theilgard, N. (2018). Intra-individual Response Variability as an Indicator of Insufficient Effort Responding: Comparison to Other Indicators and Relationships with Individual Differences. *Journal of Business and Psychology*, 33(1), 105–121.

Gao, Z., Cheng, L., Li, J., Chen, Q., & Hao, N. (2022). The dark side of creativity: Neural correlates of malevolent creative idea generation. *Neuropsychologia*, *167*, 108164.

Hao, N., Tang, M., Yang, J., Wang, Q., & Runco, M.A. (2016). A new tool to measure malevolent creativity: The malevolent creativity behavior scale. *Frontiers in Psychology*, *7*, 682.

Harris, D.J., Reiter-Palmon, R., & Kaufman, J.C. (2013). The effect of emotional intelligence and task type on malevolent creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(3), 237.

Harris, D.J., & Reiter-Palmon, R. (2015). Fast and furious: The influence of implicit aggression, premeditation, and provoking situations on malevolent creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(1), 54–64

Jonason, P.K., Foster, J.D., Egorova, M.S., Parshikova, O., Csathó, Á., Oshio, A., & Gouveia, V.V. (2017). The dark triad traits from a life history perspective in six countries. *Frontiers in Psychology*, *8*, 1476.

Kapoor, H., & Kaufman, J.C. (2022). The evil within: The AMORAL model of dark creativity. *Theory & Psychology*, *32*(3), 467–490.

Kapoor, H., & Kaufman, J.C. (2022). Unbound: The relationship among creativity, moral foundations, and dark personality. *The Journal of Creative Behavior*, *56*(2), 182–193.

Keem, S., Shalley, C.E., Kim, E., & Jeong, I. (2018). Are creative individuals bad apples? A dual pathway model of unethical behavior. *Journal of Applied Psychology*, *103*(4), 416.

Lee, S.A. & Dow, G.T. (2011). Malevolent creativity: Does personality influence malicious divergent thinking? *Creativity Research Journal*, *23*(2), 73–82.

Lubart, T., Zenasni, F., & Barbot, B. (2013). Creative potential and its measurement. *International Journal for Talent Development and Creativity*, *1*(2), 41–51.

Mastria, S., Agnoli, S., Zanon, M., Lubart, T., & Corazza, G.E. (2018). Creative brain, creative mind, creative person. In Kapoula, Z., Renoult, j., Volle, E., & Andreatta, M. (Eds.), *Exploring Transdisciplinarity in Art and Science* (pp. 3–29). Springer. https://doi.org/10.1007/978-3-319-76054-4_1

Moore, C. (2015). Moral disengagement. *Current Opinion in Psychology*, *6*, 199–204.

Nelson, J., & Botella, M. (2017). The multivariate approach and design of the creative process. *Creativity, design thinking and interdisciplinarity*, 101–113

Perchtold-Stefan, C.M., Fink, A., Rominger, C., & Papousek, I. (2020). Motivational factors in the typical display of humor and creative potential: The case of malevolent creativity. *Frontiers in Psychology*, *11*, 543948.

Perchtold-Stefan, C.M., Fink, A., Rominger, C., & Papousek, I. (2021). Creative, antagonistic, and angry? Exploring the roots of malevolent creativity with a real-world idea generation task. *The Journal of Creative Behavior*, *55*(3), 710–722.

Perchtold-Stefan, C.M., Fink, A., Rominger, C., & Papousek, I. (2022). Social exclusion increases antisocial tendencies: Evidence from retaliatory ideation in a malevolent creativity task. Psychology of Aesthetics, Creativity, and the Arts, 18(6), 1014–1025. https://doi.org/10.1037/aca0000500

Runco, M.A., Plucker, J.A., & Lim, W. (2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal*, *13*(3–4), 393–400.

Runco, M.A., Illies, J.J., & Eisenman, R. (2005). Creativity, originality, and appropriateness: What do explicit instructions tell us about their relationships? *The Journal of Creative Behavior*, *39*(2), 137–148.

Runco, M.A. & Jaeger, G.J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96.

Runco M.A. & Pagnani A.R. (2011). Psychological research on creativity. In *The Routledge international handbook of creative learning* (pp. 63–71). Routledge.

Shi, Z., Zhou, Z., Tian, L., Zhu, Y., Liu, C., & Xu, L. (2023). What causes malevolent creative people to engage in malevolent behaviors? Mediating role of moral disengagement and moderating effects of conscience. *Thinking Skills and Creativity*, *40*, 101329.

Silvia, P.J., Winterstein, B.P., Willse, J.T., Barona, C.M., Cram, J.T., Hess, K.I., … & Richard, C.A. (2008). Assessing creativity with divergent thinking tasks: exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 68.

Simonton, D.K. (2012). Taking the US Patent Office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity Research Journal*, *24*(2–3), 97–106.

Stoet, G. (2010). PsyToolkit—A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31.

Tillman, C.J., Gonzalez, K., Whitman, M.V., Crawford, W.S., & Hood, A.C. (2018). A multi-functional view of moral disengagement: Exploring the effects of learning the consequences. *Frontiers in Psychology*, *8*, 2286.

Wang, L. (2019). Creativity as a pragmatic moral tool. *Journal of Business Research*, *96*, 1–13.

Yentes, R.D., & Wilhelm, F. (2018). careless: Procedures for computing indices of careless responding. R packages version 1.2.2.

Zhao, J., Xu, X., & Pang, W. (2022). When do creative people engage in malevolent behaviors? The moderating role of moral reasoning. *Personality and Individual Differences*, *186*, 111386.