



Positive downstream indirect reciprocity scale (PoDIRS–6): Construction and psychometric characteristics

Małgorzata Szcześniak¹ · Agata H. Świątek¹ · Maria A. Świątek² · Wojciech Rodzeń¹

Published online: 24 July 2020

© The Author(s) 2020

Abstract

Downstream indirect reciprocity (DIR) is a behavior taking the form of a reaction to an individual's kindness or reluctance towards a third party. The literature shows that the concept of DIR may be understood in many different systems of assessing an individual's social exchange, retributive justice, religious belief systems, rudimentary moral systems, and general philosophical treatment, as well as from a natural selection and evolutionary approach. Given the importance of an empirically based examination of DIR, the aim of the current research carried out through Studies 1–5 was fourfold: (a) develop a reliable and psychometrically sound Downstream Indirect Reciprocity Scale (DIRS); (b) establish and examine the factor structure of the DIRS and its statistical properties, using exploratory factor analysis (EFA) (Study 1); (c) assess the relationship between the observed measures and latency factor of DIR through confirmatory factor analysis (CFA) (Studies 2–5); (d) measure the internal consistency and nomological validity (Studies 2–5). Although the first assumption consisted in constructing a questionnaire that would measure both positive and negative aspects of downstream reciprocity, the outcomes of the EFA and CFA confirmed the final version of the scale that assesses only the positive dimension of DIR (Positive Downstream Indirect Reciprocity Scale; PoDIRS-6). In fact, the EFA showed the one factor structure of the new measure, and the findings of the CFAs indicated that it meets the criteria for good fit. All of the analyses conducted defined a preliminary nomological network of convergent constructs (gratitude, life satisfaction, religiosity, and moral concerns). The PoDIRS-6 is the first scale designed to assess a set of ideas that are expressed in the belief that an individual who has done something good might get help from other people in the future. It is encouraged that a questionnaire be developed which will measure the belief that human actions can be punished or reproved when they are negative and morally bad (Negative Downstream Indirect Reciprocity Scale; NeDIRS).

Keywords Positive downstream indirect reciprocity · Gratitude · Life satisfaction · Religiosity · Morality · Psychological assessment

The events of everyday life indicate that everything related to human existence—even if we are not fully aware of it—is connected not only with direct reciprocity (you-I, I-you) but also with downstream indirect reciprocity (DIR; you-other, I-you) (Nowak and Highfield 2011). In its simplest form, DIR follows the pattern of “I scratch your back and someone else will scratch mine” (Nowak and Sigmund 2005). It is based on the principle that the generosity or wickedness we have

displayed towards others in the past will return to us in the form of kindness or disrespect from third parties who were not involved in the original interaction (Szcześniak 2018). Although the literature related to DIR has been reviewed, we are unaware of any published studies that have directly assessed the topic of DIR within the field of psychology and have presented a measurement method regarding this concept. There are only a few questionnaires that measure the direct type of reciprocity: The Personal Norm of Reciprocity, which assesses individual differences in the internalized norm of direct reciprocity (Perugini et al. 2002), and The Norm of Reciprocity Scale (Wu et al. 2006).

Given the significance of downstream reciprocity in human life, the purpose of the current study was to offer a preliminary theoretical analysis of DIR, and to present the development of a reliable and psychometrically sound scale assessing DIR. In the introductory part of the paper, we provide an overview of

This research was supported by a Grant from the National Science Centre CN No. 2017/01/X/HS6/00900.

✉ Małgorzata Szcześniak
malgorzata.szczeniak@whus.pl

¹ Institute of Psychology, University of Szczecin, Szczecin, Poland

² Pomeranian Medical University, Szczecin, Poland

previous works on DIR in the context of social exchange, retributive justice, religious belief systems, rudimentary moral systems, and general philosophical treatment, as well as DIR viewed from a natural selection and evolutionary approach. Study 1 presents the process of development of a new scale. Study 2 shows the assessment of downstream reciprocity through confirmatory factor analysis. Moreover, the analyses included in the Studies 2–5 display the results of empirical research regarding the potential correlates and predictors of DIR. Due to the lack of Polish adaptations of scales that measure belief in a just world, the concept which conceptually seems to be the closest to DIR, we suggest a preliminary set of nomological variables: gratitude, life satisfaction, religiosity, and moral concerns. Gratitude is one of the most frequently mentioned variables in the context of DIR. According to some studies (Algoe et al. 2008), it creates and strengthens social bonds among kin and nonkin. Moreover, neural researchers highlight that both a positive reputation in DIR and gratitude activate the same brain region, the dorsal precuneus (Kyeong et al. 2017; Liao 2016). Since DIR in its positive and negative forms contributes to the well-being of others, it can be hypothesized that it may be related to the life satisfaction of a reciprocating person, as well. There is some evidence (Caprara and Steca 2005) that actions that benefit others, through sharing, helping, or cooperation directly influence the life satisfaction of benefactors. Another variable that may be interconnected with DIR is religiosity. Gervais and Norenzayan (2012) report that religious people declare greater ‘public self-awareness.’ Irons (1996) observes that religion is one of the means of communicating the duties that serve as the mechanism for establishing indirect reciprocity. Finally, DIR may correlate with some measures of morality. In fact, Alexander (1986) considers systems of indirect reciprocity as moral systems since both the rewards and punishments come from individuals other than the actual recipients of the beneficence or harm.

Introduction

A review of the relevant literature reveals that there is no unanimous way of looking at DIR. One of the most pertinent perspectives implies a social exchange theory that captures reciprocal activity in a broad sense. According to this viewpoint, DIR is a behavior taking the form of a reaction to an individual’s kindness or reluctance towards a third party. Its nature is accurately illustrated by the following idiomatic expression: “I’ll scratch your back if you scratch someone else’s.” On the one hand, the discussed kind of reciprocity assumes that the person who in the past gave somebody else a hand is more likely to receive support from others in the future (Rockenbach and Milinski 2006). Conversely, somebody that did harm may meet in the future with a negative

action of somebody else. This results from the fact that individuals living in a society are subject to a cyclical exchange of good and bad things between both friends and strangers. Within these relationships, they receive rewards, take advantages, and are punished—not only in the circle of their closest friends, but also in the wider communities in which they have not helped or hurt anyone.

Another way of defining and explaining DIR is by the mechanism of natural selection, as indirect reciprocity may contribute to helpful behaviors towards people with good reputations (Nowak and Sigmund 2005). Scientists, by using models of mathematical calculations, noted that the support provided to an observed person who helps somebody else is greater when the observer’s knowledge about the helper is large enough (Yoeli et al. 2013). This knowledge may come from one’s own observations or from information obtained from other people (gossip). Studies on direct reciprocity (Burger et al. 1997; Ma et al. 2017) indicate that people are frequently kind to those who are friendly towards them. Similarly, the relationship can be extended onto a third party. People treat those who are friendly towards others with greater kindness (Rockenbach and Milinski 2006).

DIR can also be considered within the wider context of an evolutionary approach. For example, Dawkins (1976) proposes an alternative attempt to explain why people help those who support others. The author claims that supporting a person who has earlier helped somebody else may be justified by the ‘green-beard’ effect. According to this theory, the owner of a specific gene, for example connected with the tendency to help others, may recognize their own characteristic feature in other people regardless of the actual relationship between them. Depending on whether the observed person has such a feature, the carrier of the attribute of altruism will behave appropriately towards them. In the case of an action beneficial to another person, they will behave similarly towards the helper. In the case of a hostile behavior, they are likely to react in the same way. The decision is based on the ability to identify typical attributes of the altruistic gene in other people, to help them. An individual not manifesting the desired characteristic features will be deprived of support, as they do not provide it. Currently, we do not have data that would allow us to verify whether DIR is simply an expression of the green-beard gene or not. We postulate to discuss DIR as a potential phenotypic expression of a selfish gene.

Negative downstream reciprocity applied in the form of a castigation for injustice or harm caused to other people can be explained by dynamics called “altruistic punishment” (de Quervain et al. 2004). Its mechanism consists of bringing to justice the person behaving in a selfish, antisocial, exploitative or non-cooperative way. The aim of the punishment is to cause a change in their behavior in the future. The individual doing the punishing does it for the sake of others, without any real benefit for themselves (Karbowksi 2011). In empirical

studies, Fehr and Fischbacher (2003) noted that more than $\frac{2}{3}$ of respondents not participating in the exchange of good deeds, while observing behaviors towards other members of the group that were inconsistent with social norms, decided to punish the dishonest players despite the fact that the abuse did not affect them personally. Thus, altruistic punishment constitutes a key mechanism underlying the development and maintenance of cooperativeness (Niesta Kayser et al. 2011).

In the context of altruistic punishment, the occurrence of DIR can be also connected with the belief in a just world and the general idea that people receive exactly what they deserve (Kulow and Kramer 2016). Although some researchers maintain that the motivation for engaging in altruistic punishment may be related to the expectation of benefits deriving from the future interaction partners of the punisher (Fehr and Fischbacher 2003; de Quervain et al. 2004), other investigators suggest that the desire to punish may be driven by individuals' personal reactions to injustice (de Quervain et al. 2004; Kulow and Kramer 2016). For instance, Niesta Kayser et al. (2011) found that outraged observers of unfair behavior risked punishing or stopping a transgressor despite not being personally affected by the injustice. Belief in a just world expresses itself in the credence that our previous deeds result in what is actually experienced: "I deserved it" or "I am the one to blame" for my current situation. Success results from hard work, and failure equates to a lack of it. People are motivated to maintain their belief in a just world as it gives them stability and meaning in life. In this sense, belief in the existence of a just world is only a tendency to give meaning to the events of the "righteous." People assume that the world is fair, which makes them interpret events according to the principle of self-fulfilling prophecy: "someone deserved something" or "it had to end like this." However, the DIR concept goes beyond the cognitive aspect of beliefs justifying previous events, and points to the occurrence (having the nature of cause-and-effect) of indirectly rewarding and punishing behaviors within interpersonal relations. Therefore, the perception that the world is fair should rather be seen as an explanatory mechanism for DIR, and not as an alternative or an identical concept. The mechanism of altruistic punishment is an element of a range of behaviors that fall within the scope of the mechanism of DIR. According to the concept of altruistic punishment, an individual who punishes members of the community for their lack of an orientation to cooperate, does it for the good of others and does not receive from this intervention any direct benefit apart from the security that each member maintains the rules of behavior, consistent with the initially adopted standards. In the case of DIR mechanisms, it should be emphasized that they are not limited to the condition of altruistic punishment. The DIR concept places altruistic help and altruistic punishment on a common axis of behavior, which is motivated internally.

Apart from an evolutionary framework, DIR may itself be impacted by spirituality and religion (Dai et al. 2018). Several

authors (Norenzayan and Shariff 2008; Bennett and Einolf 2017) underline that the sacred texts of all major religions play a remarkably important role in promoting the prosocial norms that encourage aiding strangers, often at a personal cost. Although studies concerning DIR are connected mainly to the concept of karma, which represents a form of spirituality, and in traditional Hindu religions is thought to be part of the universal and cosmic law of cause-and-effect (McClelland 2010), the idea of indirect reciprocity is present in the content of Western religious teaching, as well: "For in the same way you judge others, you will be judged, and with the measure you use, it will be measured to you," Matthew 7.2, which means that we reap what we sow. Moreover, Blogowska and Saroglou (2011) report that religiosity is related to: prosocial values, moral behaviors, rituals, need for an ordered universe, and just-world beliefs. Religious people tend to undertake prosocial behaviors towards those who are in need but not towards observed targets who threaten traditional values. Similarly, Stavrova and Siegers (2014) note that religious individuals show prosocial actions as their Divinity awards kindness, and turn hostile and vengeful when their God approves aggression and revenge. Bushman et al. (2007) explain such behavioral tendencies through identification and justification of increasing aggression. Although the concept of karma and of DIR share some similarities, as both behaviors can be initiated by an individual because of positive or negative motives, they are different. Karma is a planned action and long-term orientation but does not refer directly to the result of the action itself. In karma, the most important role is played by the intention, not the consequence. Unlike karma, DIR-specific behaviors are effect-oriented (helping or harming the other person) to achieve a specific outcome. Moreover, the concept of karma is deprived of fatalism. For example, the negative effects of one's own actions, which might bring harmful consequences in the present or in a future incarnation (Kopalle et al. 2010), can be overcome by blurring them with actions that bear the marks of good. Instead, the DIR concept assumes the "inevitability" of the return of a good deed done, and the same being true that a person's evil will return to them. In DIR, good and bad acts do not add up. Finally, the different schools of Hinduism differ with each other in relation to whether karma is dependent on the will of the supreme god or deity, or not. DIR does not assume the participation of deities. Mediating factors characteristic of DIR are the internal dispositions of the individual (gratitude, sense of justice) and the dynamics of the interpersonal relationships.

DIR can also be seen in the context of a moral system (Alexander 1986, 1987) since it is closely related to the development of human morality (Righi and Takács 2018). Because indirect reciprocity is grounded in reputation and judging whether a person deserves to be helped (i.e., a view on what is assessed "good" or "right") or not (i.e., a view on what is

assessed “bad” or “wrong”), DIR may be interpreted as a rudimentary form of moral system (Uchida and Sigmund 2010). Individuals evaluate others’ actions as good or bad even if they are not immediately influenced by them. Righi and Takács (2018) point out that a conceptualization of indirect reciprocity can take different forms: “Give and you shall be given,” “Pay back the community the help you have received,” “Do unto others as you would have others do unto you,” “Treat others as you would be treated.” Consequently, indirect reciprocity is related to the Golden Rule, as the former drives trustworthiness and the latter is a means to judge trustworthy behavior (Boser 2014). Within this framework, downstream reciprocity is undoubtedly related to gratitude. When a person witnesses somebody else maintain the highest moral standards through helping another, often beyond the ‘call of duty’ (Ma et al. 2017), they possibly will experience moral elevation that, in turn, may translate into gratitude towards the helper. As Pohling et al. (2017) observe, the moral emotion of gratitude is not only a response to perceiving virtuous deeds of others but promotes downstream reciprocity, as well.

Finally, Hoffman et al. (2015, p. 1730) find a negative form of indirect reciprocity related to the Categorical Imperative by Kant: “Act in such a way that you treat humanity (...) never merely as a means to an end, but always at the same time as an end.” Daily life experience shows that people often feel moral disgust towards those who use or manipulate others because they are likely to mistreat their colleagues when convenient, even if, presently, the relationship is mutually beneficial.

Moreover, different empirical studies (Wenzel and Okimoto 2016; Osgood 2017) show that assessment rules implementing punitive justice—to assist those who help the good, and to not help those who do not assist the good—hark back to, among other sources, Immanuel Kant’s rule “choose that action which would, if also taken by similarly motivated others, result in a good outcome” (Hoffman et al. 2015).

On the basis of the theoretical premises discussed so far, the construct of DIR, assumed in the development of a short, self-reported scale, is grounded in a rudimentary moral system and gratitude, the religiously anchored belief system, and retributive justice as its secular equivalent. Accordingly, DIR reflects a belief that human actions can be rewarded and compensated in the future by a third person or fate when they are positive and morally good, and can be punished or reproved when they are negative and morally bad. In the proposed definition of the construct of DIR, some explicit aspects of human beliefs are present. Firstly, the description differentiates positive and negative forms of DIR. Secondly, there is a clear reference to a “third” agent who awards a benefactor. Thirdly, there is a temporal dimension of DIR. Indirect reciprocity can be carried out in the future. Implicitly, the future perspective can be near or far in time. The reciprocal response depends on the circumstances and on the personal characteristics of the “third” agent. Although the set of items presented in Table 1 conflates both beliefs and behaviors related to DIR, the main goal of the new scale was to measure beliefs rather than actions since the behavioral component of DIR would be

Table 1 Output item pool of 20 statements after experts’ assessment

1	Everything we give others in any form, will come back to us in time.
2	In helping others, we help ourselves, for whatever good we give out completes the circle and comes back to us.
3	It’s easy for me to help others, because I often get support myself.
4	All good and bad things will come back in the future to the person who began them.
5	Hearing that someone is doing good things to others, I want to support this person.
6	I easily forgive people because I have experienced forgiveness many times.
7	When I see a person who is doing good to others, I feel like being good in the same way.
8	It is easy to be nice to a person who is kind to others.
9	Life is like an echo, we get back what we give.
10	The more we give, the more we are likely to get.
11	When someone does something bad to someone else, they deserve to be punished.
12	When we help somebody, somebody else will help us.
13	Whatever I do, it will come back one way or another.
14	It is easy to be kind to a person who behaves in a pleasant way towards other people.
15	It is worth being good towards others because it will come back to us sooner or later.
16	When I see a person doing bad things to others, I would like to punish them for it.
17	Bad things will come back to those who do harm to others.
18	Good done to somebody else comes back all of a sudden, sometimes even stronger.
19	Good done to others always comes back.
20	Bad things come back like waves; when they are pushed off the shore, they come back twice as strong.

better measured experimentally. From an empirical point of view, we opted for a single-factor structure of DIR or at a maximum, for a two-factor solution, taking into account primarily the beliefs around downstream reciprocity. The construction of a multidimensional questionnaire can be advanced after validating a short questionnaire that measures one of the aspects of DIR.

Overview of the Present Research

Given the importance of empirically based examination of downstream indirect reciprocity, the aim of the current research carried out through Studies 1–5 was fourfold: (a) develop a reliable and psychometrically sound scale to measure some aspects of DIR; (b) establish and examine the factor structure of the new scale and its statistical properties, using exploratory factor analysis (EFA) (Study 1); (c) assess the relationship between the observed measures and the latency factor of DIR through confirmatory factor analysis (CFA) (Studies 2–5); and (d) assess the internal consistency and nomological validity (Peter 1981) (Studies 2–5).

With reference to validity based on relationships with other variables, we assumed that some psychological factors may enter into relations with DIR. Based on the theoretical prerequisites in the literature, gratitude, life satisfaction, religiosity, and moral concerns have been selected to develop a preliminary nomological network of interrelationships between indirect reciprocity in its downstream form (Cronbach and Meehl 1955). These variables can be considered as convergent components within a hypothesized network. We expected that they might concur and relate reasonably well with DIR, as they measure rather similar constructs. It is important to note that a belief in a just world would be an important variable to include in the nomological network of DIR since it was found to moderate the nature of a reciprocal response to an unsolicited gift (Edlund et al. 2007). However, there is no Polish adaptation of any questionnaire regarding belief in a just world. Therefore, we included the Moral Foundations Questionnaire (MFQ) with Fairness/Cheating as one of the closest factors in meaning to belief in a just world.

More precisely, in Study 2, besides CFA, we explored the association of the new scale with gratitude, which is usually seen as a social emotion or personal predisposition motivating people to undertake prosocial actions (McCullough et al. 2001), and with life satisfaction. Although most scientists perceive gratitude as a motivator of upstream indirect reciprocity (Nowak and Sigmund 1998a, 1998b; McCullough et al. 2001; Nowak and Sigmund 2005; Nowak and Roch 2007; Chang et al. 2012; Ma et al. 2017), there are grounds to assume that it also plays an important role in its downstream form. Chang et al. (2012), for example, note that gratitude leads to three forms of reciprocity: direct, upstream, and downstream. In turn, Ma et al. (2017)

emphasize that the relation between gratitude and prosocial behaviors is even stronger in the case of downstream than upstream reciprocity. If person A helps individual B, being guided not only by their sense of duty, but also reaching beyond it, it may boost moral observer C, who will be grateful and willing to help person A. McAleer (2016) assumes third-party gratitude to be the observer's reaction to the kindness and respect shown by the observed person to a stranger. The above-listed intuitions and experimental studies concerning gratitude as a motivating element to prosocial behavior allow us to assume that H1: Gratitude correlates positively with the sense of DIR. We used invariance analysis to assess the psychometric equivalence of DIRS's factor structure across genders through the combination of configural, metric, and scalar invariance, considered necessary to compare scores across groups (Milfont and Fischer 2010). The rationale behind this choice was based mainly on the previous studies which showed differences in the experience and expression of emotions between men and women (Kashdan et al. 2009). The meta-analysis investigation of gender differences in the 24 VIA character strengths, based on 65 samples (Heintz et al. 2017), yielded significant differences between females and males. Women scored higher than men in gratitude, appreciation of beauty, kindness, love, and excellence. Similarly, Cox and Deck (2006) found that generosity and reciprocity were significantly higher in women than in men. However, in other studies (Dittrich 2015) men exhibited not only more trust, but also more reciprocating behavior. Therefore, some differences between women and men in DIR can be expected.

DIR can also be related positively to life satisfaction, although at first sight, they do not seem to have a lot in common. Their relationship seems to be pertinent to measure possible correlates of DIR, as Correia et al. (2009) found empirical evidence that belief in a just world, which lies at the roots of reciprocal behaviors, leads to well-being, and vice versa. Other studies reveal that this belief co-occurs with the sense of one's own happiness (Dalbert 2001; Lucas et al. 2013). Being aware that people in our society comply with norms, and if they do not comply, they deserve an appropriate punishment, strengthens the sense of trust, hope and belief in the future in individuals (Lerner 1980). The abovementioned elements constitute essential indicators of satisfaction because well-grounded trust and optimism in the context of everyday challenges contribute to a greater sense of fulfilment. Furthermore, social relationships described as equality matching (Fiske 1992), based on a balanced exchange between people, boost the sense of well-being (Kim and Kim 2003) and happiness. Therefore, in the context of the earlier findings, it can be assumed that H2: DIR will correlate positively with life satisfaction.

In Study 3 and Study 4, we conducted an analysis between DIR and religiosity, which seems to be another factor, representing the preliminary group of nomological

components bounded positively with DIR. Tullberg (2004, 2012) noticed that many moral systems indicate being an altruist as beneficial. Religions usually assure their followers that there is eternal life or reincarnation, expressing a relation model typical for DIR: When A helps B, later C (God or god) will reward every good deed in a way commensurate to their nature. This compensation is called a “metaphysical reward.” At the same time, the faithful are also warned against negative behaviors because they may result in a “metaphysical punishment.” A’s action harmful to B may cause a punishment imposed by C (God or god). Nordin (2015) adds that the belief in a divine or human warning has a communicative function as a way to increase cooperation in the community. People who do not cooperate with others or do not reward them for good deeds, even if they do not receive any benefits directly, expose themselves to stigmatization and ostracism in their communities. This leads to a sense of social isolation and loneliness. Therefore, it can be assumed that H3: Religiosity correlates positively with DIR.

In Study 5, we analyzed the relationship between DIR and moral concerns. This choice was based on the premise (Righi and Takács 2018) that indirect reciprocity is particularly pertinent to questions of human morality, and related to its development. The mechanisms of rewarding for good deeds and punishing for bad ones might coexist with five moral concerns of Moral Foundations Theory (Graham et al. 2013): 1) the adaptive challenge for caring and protecting that is characteristic of the care/harm foundation; 2) the sensitiveness to evidence of cheating and cooperation that is specific to fairness/cheating and reminiscent of the concept of justice; 3) the ability to form cohesive coalitions based on trust, obligations, and compromise that distinguishes the loyalty/betrayal domain; 4) the capacity to live in and build hierarchical social groups; respect relational chains of authority; 5) the ability to make decisions grounded not only in the sensory properties, but primarily in the virtues of temperance and chastity. Although all five components of Moral Foundations Theory theoretically relate to DIR, the second moral concern seems specifically important in the context of downstream reciprocity since fairness is a source of social obligation (Janoff-Bulman and Carnes 2013). Because of this association, we chose all five moral concerns as factors with predictive power for DIR. On the basis of these premises, we assumed that H4: All moral concerns will result in positive correlations with the sense of DIR.

Study 1

Development of Original Item Pool

In the first study, after delineating (Boateng et al. 2018) DIR as a belief that human actions can be rewarded and

compensated in the future by a third person or fate when they are positive and morally good, and can be punished or reproved when they are negative and morally bad, we generated a pool of twenty-eight statements which hypothetically were to constitute the core of the scale for the measurement of DIR. We respected the recommendation that the original pool of items should be at least twice as long as the desired final scale (Schinka et al. 2012). In respect to the form of the items, we tried to capture in a simple and unambiguous way (Schinka et al. 2012) the main characteristics of DIR: the return of good or bad deeds, and the third-party response. No reverse-scored items were included for two reasons. First, we followed one of the principal guidelines for item development consisting in the avoidance of negative formulations. Second, inverting items by using an antonymic expression can create difficulties of interpretation because the connotation of the item can alter (Suárez-Alvarez et al. 2018).

The formulation of the set of items was developed through a deductive approach (Boateng et al. 2018), based on the theoretical foundations from the literature review presented in the first part of the paper. It was an a priori phase undertaken to enhance content validity (Polit and Tatano Beck 2006) through a conceptualization of DIR. Next, the relevance of the items, their comprehensiveness and representativeness, were examined through three experts’ assessments provided by researchers familiar with the topic of indirect reciprocity (a posteriori stage) (Yaghmaie 2003). The main criterion for the experts to refine the measure was choosing only the best items (Schinka et al. 2012) relevant to the DIR construct. Based on the approach of average congruency percentage (Waltz et al. 2005), we identified which items were judged to be congruent, taking as acceptable only those which scored at least .90. In a few cases, the experts also made some modifications to the word choices. We eliminated items based on the experts’ evaluations (Reich et al. 2018), reducing the pool to 20 items. The first version of the questionnaire is presented in the Table 1. The respondents evaluated each of twenty statements by using multiple-choice answers on a 7-point Likert scale that ranges from 1 = *strongly disagree* to 7 = *strongly agree*.

Although the remaining twenty proposed statements do not mention the concept of a “third party” and do not refer directly to good or bad “reputation,” they are implied in the content of some items (i.e.: “When I help somebody, somebody else will help me” or “When someone does something bad to someone else, they deserve to be punished”). Moreover, the time horizon is also presumed and indicates the fact that the reply in the form of an award or punishment may come in an unspecified future (i.e.: “It is worth being good towards others because it will come back to us sooner or later”). Therefore, it was assumed that the new scale would help to study the idea of DIR, expressed in the belief that a person who has done something good in the past is more likely to receive support from other people in the future. On the other hand, somebody that did

harm might meet in the future with a negative behavior of somebody else.

Participants

The study comprised 264 people (76.0% women, $M_w = 26.92$; $SD = 11.07$; 24% men, $M_m = 27.30$; $SD = 11.94$; $p = 0.795$) who were approached from a variety of channels and selected by means of a non-probability convenience sampling procedure. The participants were aged between 18 and 74. The average age was approx. 25 ($M = 25.01$; $SD = 11.83$). The sample was composed of young adults (80%; 18–39 years old), the middle-aged (19%; 40–65 years old), and older adults (1%; 66–74 years old). The group of younger respondents consisted of undergraduate students from introductory psychology, medicine, and national security courses who were participating in partial fulfilment of a course research requirement. The group of older adults included professionals who had various levels of educational and work experience: teachers, economists, doctors, accountants, managers, soldiers, lawyers, IT specialists, sole traders, and entrepreneurs. All respondents were assured of the confidentiality of their information and gave informed and written consent for their participation in the study.

Procedure and Data Analysis

A preliminary data analysis was conducted to direct the successive exploratory factor analysis. First, all the DIRS items were screened for skewness and kurtosis to evaluate the normality of item distribution (Muthén and Kaplan 1985). We assumed values less than ± 2 as a normal distribution (George and Mallery 2014).

In order to control for potential common method variance (CMV), we followed the strong recommendation by Williams and colleagues (Williams and McGonagle 2016; Williams and O’Boyle 2015) to use a CFA marker variable technique as a procedural remedy for CMV. We chose self-esteem as a marker variable since this psychological construct seems theoretically unrelated to DIR and should not correlate significantly with indirect reciprocity.

Next, given the novel and exploratory character of the DIRS, we considered it appropriate to use an EFA with a one factor structure of the scale, as it was expected that the items would be similar. Following a widely cited rule of thumb (Nunnally 1978) that the subject-to-item ratio for an EFA should now be lower than 10:1, we performed the EFA with a maximum likelihood (ML) estimation (promax rotation), with a subject-to-item ratio of 13:1. We confirmed that the data were appropriate for this technique through the Kaiser-Meyer-Olkin Measure of Sampling Adequacy used as a factor retention method. However, because of the weakness of the Kaiser criterion (Wood et al. 2015), a scree plot

was chosen due to its visual nature, and because it is typically considered a good estimate of the ideal number of factors to retain (Osborne 2014). The ML procedure, as an extraction method, was selected based on the recommendation of Fabrigar et al. (1999), who argue that if the data are normally distributed, ML is the best solution. Next, we performed a parallel analysis in SPSS through the use of the rawpar.spss script developed by O’Connor (2000), which is considered advantageous over the more classical approaches (Osborne 2014) and is believed to be the only method that formally measures the probability that a factor or factors are due to a chance (Wood et al. 2015). A Monte Carlo simulation with 1000 replications on randomly generated data was run to determine the number of principal significant components to retain for further analysis (Franklin et al. 1995). The rule underlying the chosen factor/s imparts that factors corresponding to actual eigenvalues should be greater than the parallel average random eigenvalues.

The present research project and all of the following studies were approved by the Bioethics Committee of the Institute of Psychology at the University of Szczecin (KB 12/2017). All computational procedures were performed with the use of IBM SPSS statistics package version 20 and IBM SPSS AMOS 21.

Results

A review of the summary statistics showed an acceptable range for the majority of items, and an abnormal distribution in the case of four items which revealed higher and unsatisfactory values of skewness or kurtosis. Thus, items 1, 3, 8, and 14 were excluded from subsequent analyses. In addition to skewness and kurtosis, the mean, standard deviation, minimum, and maximum values are presented in Table 2.

The results of a CFA marker variable technique used through IBM SPSS AMOS 21 showed that the CMV decreased when self-esteem was included in the model. In fact, the common variance based only on the items of DIR was about 37%. After adding items of self-esteem (SES by Rosenberg), the common variance was reduced to less than 1%. Since the basic assumption of this method is that a general factor does account for a majority of the covariance among the variables (Melas et al. 2011), the outcome suggests that common method bias was not an issue for the present study.

High Kaiser-Meyer-Olkin statistics (0.908) and a significant probability level smaller than $p < 0.001$ for the Bartlett’s Test of Sphericity of all sixteen items indicated sufficient correlations to proceed with factorial analysis ($\chi^2 = 2906.332$, $df = 120$). The results confirm that the sample size was appropriate for factor analysis.

The pool of 16 remaining items was subjected to EFA with maximum likelihood (ML) estimation (promax rotation). Three components had eigenvalues greater than 1.0,

Table 2 Descriptive statistics, skewness and kurtosis

Items	Mean	SD	Min	Max	Skewness	Kurtosis
DIR1	5.88	1.28	1	7	-1.53	2.53
DIR2	5.63	1.42	1	7	-1.21	1.25
DIR3	6.03	1.01	1	7	-1.36	2.85
DIR4	5.38	1.58	1	7	-0.92	0.18
DIR5	4.93	1.70	1	7	0.61	-0.64
DIR6	3.71	1.60	1	7	0.08	-0.90
DIR7	3.77	1.79	1	7	0.02	-1.19
DIR8	5.72	1.27	1	7	-1.45	2.72
DIR9	5.28	1.56	1	7	-0.94	0.37
DIR10	5.20	1.53	1	7	-0.97	0.55
DIR11	4.88	1.74	1	7	-0.65	-0.45
DIR12	5.03	1.40	1	7	-0.75	0.53
DIR13	4.98	1.50	1	7	-0.69	0.12
DIR14	5.59	1.21	1	7	-1.47	3.07
DIR15	5.37	1.46	1	7	-1.15	1.06
DIR16	4.35	1.73	1	7	-0.36	-0.76
DIR17	4.86	1.59	1	7	-0.73	-0.04
DIR18	5.15	1.49	1	7	-0.87	0.52
DIR19	5.20	1.54	1	7	-0.94	0.44
DIR20	4.77	1.66	1	7	-0.54	-0.45

accounting between them for 60.7% of the variance, and a visual scree plot suggested that a three-factor solution might yield interpretable factors (Fig. 1).

Nevertheless, we tentatively chose to take into consideration only the first factor, as its eight items had very good

loadings (larger than 0.6) and an excellent Cronbach alpha of 0.94 (Table 3). We used a more stringent cut-off of 0.6 in order to ensure a more robust and consistent scale; a practice applied within an exploratory context, especially when short questionnaires are being developed (Altmann and Roth 2018; Jones et al. 2018; Park 2014; Shen et al. 2011). In fact, according to Tabachnick and Fidell (2007), cut-offs of 0.63 are very good. Although the second factor had three item loadings close to the value 0.63 (numbers 5, 7, and 16), this was not satisfying for several reasons. First, in terms of content, three items dealt with the personal implementation (behavior, action) of reciprocity and not with the aspect of belief proposed in the definition of DIR. Moreover, two of the three items had a positive connotation related to support for good actions, and one item indicated negative behavior that would require punishment for misdeeds. In other words, they represented two separate axes of reciprocity. Therefore, it would be more appropriate to enlarge the group of items that have a behavioral character, respectively specifying a set of negative and positive statements. Finally, the lack of a certain consistency between these items could be confirmed by a lower reliability value, which was $\alpha = 0.75$. With respect to a third potential factor, only two items met the requirement of very good loadings (numbers 17, and 20). Since two-item factors are not recommended because they may present a problem for identification, we abandoned it in our further research.

Finally, using the 16-item data set, a parallel analysis also indicated a three-factor solution for both the mean and 95th percentile eigenvalues, although factors II and III were just above the mean ($M = 2.47$ and $M = 1.27$) and 95th percentile cutoffs (1.41 and 1.32). Hence, it seems that the PA implies

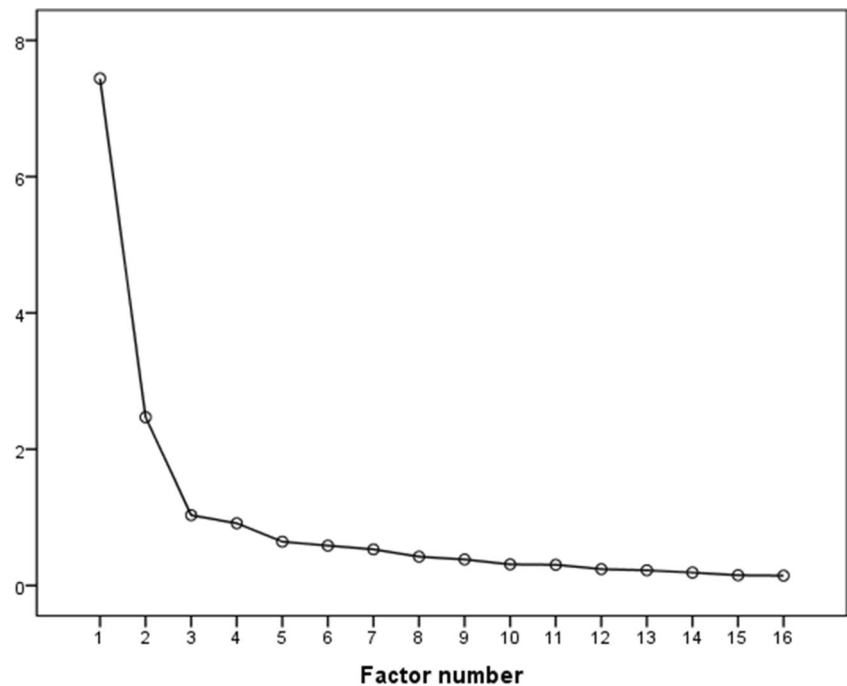
Fig. 1 Scree plot

Table 3 Promax rotation

Items	Factor		
	1	2	3
DIR2	0.71	−0.07	0.14
DIR4	0.48	0.02	0.46
DIR5	0.06	0.64	−0.10
DIR6	0.25	−0.12	−0.01
DIR7	0.08	0.92	−0.12
DIR9	0.63	0.06	0.28
DIR10	0.65	0.08	0.07
DIR11	−0.18	0.38	0.36
DIR12	0.79	−0.09	−0.08
DIR13	0.98	0.01	−0.11
DIR15	0.91	0.05	−0.06
DIR16	−0.12	0.62	0.22
DIR17	0.27	0.02	0.65
DIR18	0.85	−0.01	0.04
DIR19	0.85	−0.01	0.09
DIR20	0.35	−0.02	0.58

one clear factor and two weaker ones, which are both rather uninterpretable and unlikely to be replicated, confirming the outcomes of the scree plot.

Taking into account the results of both the scree plot and the PA, we ultimately decided to retain only the first factor (8 items: 2, 9, 10, 12, 13, 15, 18, and 19) (Table 3). We followed such a procedure because it is acknowledged that factors that have less than three variables might not explain the total variance meaningfully, and are generally viewed as undesirable and trivial (Yong and Pearce 2013).

Discussion

In Study 1, we aimed to: (a) develop a psychometrically reliable Downstream Indirect Reciprocity Scale (DIRS), (b) establish the initial factor structure of the DIRS and its statistical properties, using exploratory factor analysis (EFA). Overall, the number of factors to be retained was guided by theory (item content), examination of the Kaiser-Guttman criterion, a scree plot, and Horn's parallel analysis. Although the decision of one factor might seem too restrictive, it has its empirical justification. With regard to the PA, though relatively accurate, it is still inclined to the error of suggesting the retention of one or two more factors than are generally warranted, and of keeping poorly defined factors (Glorfeld 1995). With regard to a cut-off of 0.6, Hair et al. (2017) suggest a strong cut-off of 0.7, and other authors (Altmann and Roth 2018; Jones et al. 2018) postulate using this criterion for brief questionnaires. The interpretation of the EFA was guided by an expected a priori one-factor solution.

Study 2

Participants

The study comprised 317 people (233 women—73.5%) through a snowball recruitment method. The participants were aged between 17 and 74. The average age was 27 ($M = 26.97$; $SD = 11.27$). The sample was composed of adolescents (5%; 11–17 years old), young adults (79%; 18–39 years old), the middle-aged (15%; 40–65 years old), and the elderly (1%; 66–70 years old). Among the respondents, 37% were high school or university students, and 63% were employees or pensioners. The under-age respondents took part in the study after obtaining their carer's consent. Adolescents were allowed to participate as the instrument seems to be suitable for minors, who at this stage of life are developing strong principles of fairness, justice, and equality (Cooley et al. 2012). All other participants gave informed and written consent for their participation in the study. All the respondents were reassured of the confidentiality of their information.

Procedure and Data Analysis

As structural equation modelling requires the variables to be normally distributed, prior to the CFA, the univariate normality of all 8 DIRS items identified with the EFA was checked through examining the skewness and kurtosis at the item level.

Next, three CFAs were performed, using a covariance matrix, for different DIRS models of a single-factor solution: 8-item, 7-item, and 6-item solutions. To assess how much variance of the indicators is explained by the latent variable, the loadings of the indicators were examined. Since there is no consensus on which measures are the most suitable to evaluate the goodness of fit (Brown 2006; Byrne 2008), some of the most common fit indices were used in the estimation of the model (Brown 2006): the Chi-Square value, the Minimum Discrepancy to Degree of Freedom (CMIN/DF), the Goodness-of-Fit Index (GFI), the Tucker-Lewis Index (Tucker and Lewis 1973), the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA) and its 90% Confidence Interval, the test of the Closeness of Fit (PCLOSE), and the Hoelter critical number of 0.05 and 0.01. The thresholds for these fit statistics for allowing or rejecting the model are given in Table 4. In examining the models, we took into consideration a more conservative approach: nonsignificant Chi-Square and $RMSEA \leq 0.05$.

Because the initial model did not meet the criteria for a good fit, we conducted model modifications to the original hypothesized model to get a better fitting. Although this practice is considered controversial, some modifications are allowed (Hooper et al. 2008) to improve the results. We based our procedure on statistical evidence (modification indices) and conceptually redundant items, as the re-specification

Table 4 Summary of fit indices and minimum acceptance levels

	Fit index	Suggested value
1.	χ^2/df ratio p	There is no theoretical maximum for the χ^2/df ratio (West et al. 2012) Nonsignificant Chi-Square is desired (Bentler and Bonnett 1980)
2.	CMIN/DF	< 2 (Byrne 2008)
3.	GFI	> 0.9 (Hu and Bentler 1999)
4.	TLI	0.97 is accepted as the cut-off value (Cangur and Erkan 2015)
5.	CFI	> 0.9 (Bentler 1990) ≥ 0.95 (Hu and Bentler 1999; Cangur and Erkan 2015)
6.	RMSEA	≤ 0.06 Good fit (Hu and Bentler 1999)
	LO	0.05 ≤ < 0.08 Reasonable fit (Browne and Cudeck 1992; Byrne 2008)
	HI	0.08 ≤ < 0.1 Mediocre fit > 0.1 Poor fit (Browne and Cudeck 1992; MacCallum et al. 1996) ≤ 0.06 (Hu and Bentler 1999; Brown 2006)
7.	PCLOSE	≥ 0.05 and ≤ 1 (Browne and Cudeck 1992)
8.	CN .05	≥ 200 (Hoelter 1983; Diamantopoulos and Siguaw 2000)
	CN .01	>> 200 when sample size is above 250 (Hu and Bentler 1995)

requires a strong empirical and theoretical justification (Schreiber et al. 2006; Byrne 2008; Mueller and Hancock 2008). We presented the fit indexes and Chi-Square values of all three models in Table 6.

After assessing the fit indexes of the 6-item model for the whole sample, an invariance analysis was used to assess the psychometric equivalence of the new scale's factor structure across genders. The rationale behind this analysis was based on two important premises. Firstly, the present research group was predominantly female. Secondly, we assumed that gender could be a factor that has an effect on DIR. In fact, researchers who have examined gender differences in the belief in a just world have found mixed results (Cohn and Modecki 2007; Reich and Wang 2015). Some results based on a meta-analytic review suggest that there is no significant difference in this respect between males and females (Harper et al. 1990; O'Connor et al. 1996). However, some other studies (Karadağ and Akgun 2016) indicate that such a difference exists. Male students had stronger general and personal just-world beliefs than females. The analysis was conducted through the combination of configural, metric, and scalar invariance, considered necessary to compare scores across groups (Milfont and Fischer 2010; Reich and Wang 2015). Similarly, gratitude and religiosity research displays sex differences. A series of studies predominantly shows that women obtain higher scores than men on the indexes of gratefulness (Kashdan et al. 2009), thankfulness (Kaczmarek et al. 2015), religiousness (Francis 1992; Roth and Kroll 2007), and spirituality (Flannelly and Galek 2006).

Finally, the nomological validity of the scale (Appendix 1) was tested with the use of the Gratitude Questionnaire (GQ-6; McCullough et al. 2001; Kossakowska and Kwiatek 2014). The correlation was controlled by life satisfaction (SWLS; Juczyński 2009) to verify if and how being satisfied with one's

life can influence the relationship between downstream reciprocity and gratitude. A partial correlation was used to verify if the observed relationship between both variables might be distorted by the influence of satisfaction. Moreover, we used a linear regression model to check if gratitude and satisfaction were predictors of DIR, and if there were some outliers in the sample. We computed Mahalanobis' distance, using the chi-square distribution with a very conservative probability estimate for a case being an outlier ($p < 0.001$), and Cook's distance.

The rationale for measuring downstream reciprocity in relationship with gratitude has been previously explained as the association between being grateful and prosocial behaviors seems to be strong in the case of DIR (Ma et al. 2017). Yet, downstream reciprocity can also be related positively to life satisfaction, although at first sight they do not seem to have a lot in common. Correia et al. (2009) found empirical evidence that belief in a just world, which lies at the root of reciprocal behaviors, leads to well-being and vice versa. Other studies reveal that this belief co-occurs with the sense of one's own happiness (Dalbert 2001; Lucas et al. 2013). Being aware that people in our society fulfil norms, and if they do not comply, they deserve an appropriate punishment, strengthens an individual's sense of trust, hope and belief in the future (Lerner 1980). Therefore, we decided to investigate not only the relationship between DIR and gratitude, but also the role of satisfaction within this relationship.

Measures

To test the nomological validity of the new scale, we applied the following measures in Study 2: the Gratitude Questionnaire (McCullough et al. 2001; Kossakowska and Kwiatek 2014),

and the Satisfaction with Life Scale (Diener et al. 1985; Juczyński 2009).

The Gratitude Questionnaire (GQ-6) was used to examine gratitude understood as a predisposition. It consists of six statements on which it was necessary to take a position using a seven-point Likert scale (from 1—“I strongly disagree” to 7—“I strongly agree”). The higher the result, the higher the gratitude level (Kossakowska and Kwiatek 2014). The reliability of the GQ-6 obtained in the current study was slightly higher (Cronbach $\alpha = 0.73$) than the results of the psychometric parameters gained in the original version (Cronbach $\alpha = 0.71$) (Kossakowska and Kwiatek 2014).

The Satisfaction with Life Scale (SWLS) developed by Diener et al. (1985) was used to assess the sense of satisfaction with life. The tool consists of 5 statements assessed on a seven-point scale (from 1—“I strongly disagree” to 7—“I strongly agree”). The scores after summing up give the overall result which indicates the level of satisfaction with one’s own life. The scope of results ranges from 5 to 35. The higher the result, the greater the satisfaction with their lives that the respondents show. In our study, the questionnaire was assessed as satisfactorily reliable (Cronbach $\alpha = 0.81$).

Results

The results showed that all of the skewness and kurtosis values were less than ± 2 (Bowen and Guo 2012), thus confirming that the items can be considered to be normally distributed (Table 5).

The eight-item model suggested a rather poor fit (Table 6) with loadings of .78, .77, .71, .73, .88, .86, .85, and .87, respectively. On the basis of high modification indices (26.134) and item content overlap (item 12—“When I help somebody, somebody else will help me” and item 13—“Whatever I do, it will come back one way or another”), we proceeded into exploratory mode (Byrne 2008) and tested a seven-item model, removing item 13. Although this change did not result in an adequately fitting model, it provided a better fit than the first one, with the loadings of .77, .69, .73, .89, .86, .85, and .87. Using the criteria of high modification indices (25.608) and its

content resemblance one more time, we decided to remove item 10 (“The more we give, the more we are likely to get”) as it corresponded to the content of item 18 (“Good done to somebody else comes back all of a sudden, sometimes even stronger”). After this adjustment, the six-model scale fully met the criteria for a good model fit (Table 6) with loadings of .81, .78, .74, .88, .88, and .91 (Fig. 2).

Therefore, it seems that the modification of the confirmatory factor model did not lead to a model that is substantially different from the theoretical model originally hypothesized on the basis of the earlier EFA, and could be verified in the following two studies. The internal consistency of the PoDIRS-6 (Positive Downstream Indirect Reciprocity Scale) was estimated by Cronbach’s Alpha, which was equal to 0.93.

The results of configural invariance across genders without constraints showed a good model fit to the data: $\chi^2(18) = 16.647, p = 0.548$; CFI = 1.000; GFI = 0.982; RMSEA = 0.000 [LO 90 = 0.000, HI 90 = 0.046]; PCLOSE = 0.967; Hoelter = 544 (0.05), and Hoelter = 656 (0.01). This outcome indicates that the pattern of factor loadings of six items on the latent DIR construct holds across both groups and is the same (“invariant”) for men as it is for women. This means that respondents from both gender groups conceptualized the construct and individual statements, underlying latent factor, in the same way (Jorgensen et al. 2018; Putnick and Bornstein 2016; Reich and Wang 2015; van de Schoot et al. 2012). The next level of invariance tested was metric invariance, which requires the constraint of factor loadings to be equivalent across independent samples (Jorgensen et al. 2018). The model passed this level of analysis: $\chi^2(23) = 20.319, p = 0.623$; CFI = 1.000; GFI = 0.978; RMSEA = 0.000 [LO 90 = 0.000, HI 90 = 0.040]; PCLOSE = 0.987; Hoelter = 543 (0.05), and Hoelter = 643 (0.01). Finally, a test for scalar invariance was conducted by constraining item intercepts to be equivalent across the female and male groups. This model fit well: $\chi^2(30) = 31.958, p = 0.369$; CFI = 0.998; GFI = 0.978; RMSEA = 0.014 [LO 90 = 0.000, HI 90 = 0.046]; PCLOSE = 0.974; Hoelter = 430 (0.05), and Hoelter = 500 (0.01). All results indicate that the observed scores are related

Table 5 Descriptive statistics, skewness and kurtosis, reliability coefficients if item deleted

Items	Mean	SD	Min	Max	Skewness	Kurtosis	α if item deleted
DIR2	5.55	1.13	1	7	-1.08	0.86	0.93
DIR9	5.18	1.60	1	7	-0.83	0.06	0.93
DIR10	5.11	1.57	1	7	-0.90	0.31	0.93
DIR12	4.92	1.40	1	7	-0.72	0.52	0.93
DIR13	4.92	1.50	1	7	-0.64	0.03	0.92
DIR15	5.29	1.45	1	7	-1.00	0.70	0.92
DIR18	5.10	1.51	1	7	-0.79	0.27	0.93
DIR19	5.11	1.57	1	7	-0.82	0.06	0.92

Table 6 Goodness-of-Fit Indices for three models

Model	χ^2 *	d.f.*	P	CMIN	GFI	TLI	CFI	RMSEA [LO; HI]	PCLOSE	CN .05	CN .01
8 item	92.232	20	.000	4.612	.932	.948	.963	.107 [.08; .12]	.000	108	129
7 item	45.963	14	.000	3.283	.961	.971	.980	.085 [.06; .11]	.017	163	201
6 item	14.045	9	.121	1.561	.986	.993	.996	.042 [.00; .08]	.573	381	488

χ^2 */d.f.: relative Chi-Square; Minimum Discrepancy to Degree of Freedom (CMIN/DF), Goodness-of-Fit Index (GFI), Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA) and its 90% Confidence Interval [LO; HI], Closeness of Fit (PCLOSE), and Hoelter critical number 0.05 and 0.01

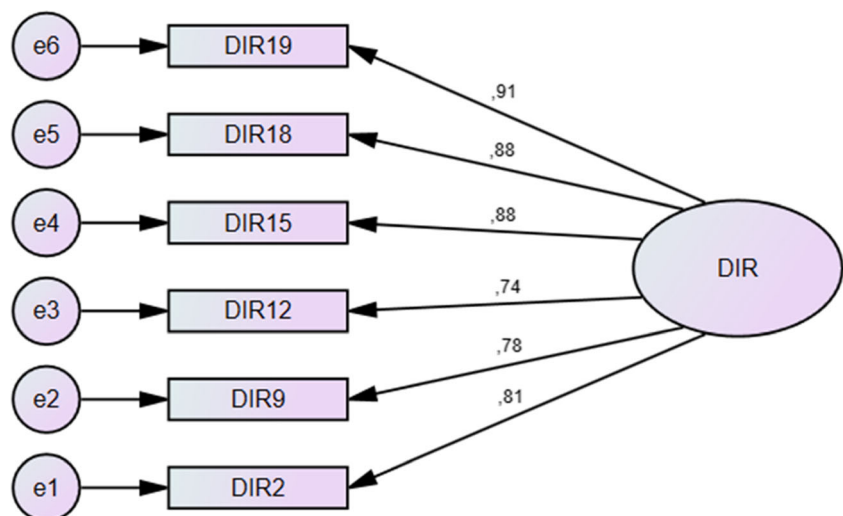
to the latent scores, and the intercepts are invariant across the female and male groups. Taking into consideration all of the outcomes, it was assumed that the measurement invariance of the PoDIRS-6 holds across sexes.

With respect to correlation, downstream reciprocity was positively associated with gratitude ($r[315] = .390, p < 0.001$), and life satisfaction ($r[315] = .235, p < 0.001$). Moreover, gratitude positively correlated with life satisfaction ($r[315] = .309, p < 0.001$). Next, we tested the first-order partial correlations among downstream reciprocity, gratitude, and life satisfaction. When controlling for life satisfaction, the correlation between downstream reciprocity and gratitude was reduced but remained significant ($pr[315] = .344, p = .001$). The results of a linear regression analysis showed that the DIR beliefs were predicted by both gratitude ($\beta = 0.329; t = 5.985; p < 0.001$), and life satisfaction ($\beta = 0.110; t = 2.006; p < 0.05$) with $F(316,2) = 26.109; p < 0.001$. With respect to outliers, none of 317 cases was identified as a possible multivariate outlier. In fact, the lowest value of probability was 0.00226. Cook's value (between 0.000 and 0.072) was under the point at which the researcher should be concerned (less than 1).

Discussion

The aim of Study 2 was to measure the internal consistency and nomological validity of the new scale. Because the initial model did not meet the criteria for a good fit, we conducted model modifications to the original hypothesized model in order to get a better fitting. According to Anderson and Gerbing (1988), initially specified measurement models regularly fail to provide a suitable fit and a series of respecifications is required based on a combination of statistical and theoretical considerations. In this sense, the CFA in the present study was used in an exploratory way (Brown 2006) as we inspected modification indices and removed two items overlapping in terms of content. We took into consideration Schmitt's (2011) remark that modification post hoc can be considered by researchers, though carefully, to further explore poor fitting models. We also took account of Anderson and Gerbing's (1988) recommendation that, after achieving an acceptable fit within the respecification procedure, the next step is to validate the final model through a CFA within the context of another sample. We did this in Studies 3–

Fig. 2 Measurement model of final DIR



5. At this point, it is important to note some key points with regard to the new measure. The final version of the PoDIRS-6 (Positive Downstream Indirect Reciprocity) questionnaire, as a result of statistical calculations, constrained us to narrow a definition of only its positive aspects. In fact, the single factor contains only positive items regarding downstream indirect reciprocity. In this new form, adapted to the psychometric construction of the scale, positive DIR reflects the belief that human actions can be rewarded and compensated in the future by a third person or fate when they are positive and morally good. Such a description highlights the positive dimension of DIR, and in this sense is a reasonable way to capture only this facet of indirect reciprocity. At the same time, DIR in its negative configuration (“the belief that human actions can be punished or reproved when they are negative and morally bad”) may be elaborated in the future and called NeDIRS (Negative Downstream Indirect Reciprocity), being the negative equivalent of the positive version: PoDIRS-6. This is due to the fact that the initial question set did not adequately grasp the features of NeDIRS, and a focused effort will be needed to identify additional items that measure the negative characteristics of downstream indirect reciprocity. Finally, there is a possibility that the action components of positive and negative forms of downstream reciprocity may also need to be separately measured.

In accordance with the goal of Study 2, gratitude correlated positively with PoDIRS-6. This finding is in line with studies on moral elevation (Pohling et al. 2017). Since moral elevation and gratitude are responses to witnessing virtuous deeds of others, elevation may lead people to appreciate and emulate observed acts of kindness. Being thankful for a gesture of benevolence towards someone else may, in the future, express itself through another act of generosity in the name of the belief that an observed person who did something good in the past deserves the same kindness and respect. Because the effect of the associations between positive DIR and gratitude was medium by Cohen’s criteria, we can assume that this relationship is realistic and appropriate. In fact, Hemphill (2003), referring to Cohen (1988), who proposed similar empirical guidelines for interpreting the magnitude of correlation coefficients (those of 0.10 are “small,” of 0.30 are “medium,” and of 0.50 are “large”), remarks that the value used to denote a large correlation occurs rather occasionally in many key psychological research studies.

Life satisfaction was associated positively with PoDIRS-6. This outcome confirmed previous studies. For example, research drawing a conceptual framework from social exchange theory shows that fairness (Cate et al. 1982), social norms, and a sense of congruence (Suh et al. 1998) are strong predictors of relational satisfaction. In other studies, belief in a just world led to well-being (Correia et al. 2009) and co-occurred with the sense of one’s own happiness (Dalbert 2001; Lucas et al. 2013).

Both gratitude and life satisfaction were predictors of PoDIRS-6. With respect to gratitude as a stronger predictor, the outcomes seem to confirm that gratitude may promote prosocial (Peng et al. 2019) and honest behavior (DeSteno et al. 2019). In fact, in previous studies, grateful individuals were found to have a higher propensity to build social bonds (Algoe et al. 2008), foster cooperation (Vayness et al. 2019), and maintain norms in an effort to stabilize group harmony (Ng et al. 2017). Gratitude facilitated the increase and maintenance of mutual direct and indirect relationships (Balconi et al. 2020) as well. With regard to satisfaction as a weaker but still significant antecedent of positive DIR, there has been some evidence that people who experienced positive moods were also likely to engage in extra-role behaviors, although there were not formally required to do so (George 1991).

Study 3

Participants

The study comprised 296 people (212 women—71.6%). The participants were aged between 16 and 86. The average age was 25 ($M = 25.34$; $SD = 9.95$). The sample was composed of adolescents (2%; 16–17 years old), young adults (88%; 18–39 years old), the middle-aged (8%; 40–65 years old), and the elderly (2%; 66–86 years old). Among the respondents, 13% were high school students, 60% were university students, and 27% were employees or pensioners. Parental consent was obtained for all respondents below 18. All participants were reassured that their participation in the study would be confidential. Participants were recruited through similar strategies to those described for Studies 1 and 2, this is, via snowball sampling (different social networks and local organizations).

Procedure and Data Analysis

Prior to the other analyses, skewness and kurtosis were calculated to study the characteristics of the distribution of PoDIRS-6. As previously, we assumed values less than ± 2 as a normal distribution (George and Mallery 2014).

Next, a CFA was carried out on new data collected separately of the original study in which a factor structure was derived. The aim of this assessment was to investigate the extent to which the 6-item solution could be replicated in a subsequent CFA study, after modification of the original model of the questionnaire (Study 2). The rationale behind examining the PoDIRS-6 structure once again was to verify the psychometric stability (one-factor solution with good fit indices) of the new measurement method.

Measures

The accuracy of PoDIRS-6 was tested with the use of the Scale of Religious Meaning System developed by Krok (2011). Its main feature is the inherent relationship with the sanctity sphere as well as with indicative and sense-creating factors (Krok 2014). The scale comprises 20 statements, creating two main dimensions: (1) the dimension of religious orientation refers to the understanding ourselves and relationships with other people and the world, and (2) the dimension of religious sense regards the interpretation of life in terms of meaning and purpose. Factors are assessed on a seven-point Likert scale. The original Cronbach α internal compliance indices are satisfactory and their values are: α orientation scale = 0.92; α sense scale = 0.89; the entire α scale = 0.93 (Krok 2011, 2014). In this study, the Cronbach α for the entire scale was $\alpha = 0.96$, for the orientation scale: $\alpha = 0.93$, and for the sense scale: $\alpha = 0.92$.

Results

Skewness and kurtosis were within the acceptable limits of ± 2 (Table 7).

CFA estimations with no missing data confirmed a single-factor solution of PoDIRS-6 with strong loadings of .78, .81, .70, .81, .80, and .82, respectively, and a good model of fit with the following goodness of fit indicators: $\chi^2(9) = 13.94$, $p = .124$; CMIN/DF = 1.550; GFI = 0.984; TLI = 0.992; CFI = 0.995; RMSEA = 0.043 (LO 90 = 0.000; HI 90 = 0.085); PCLOSE = 0.551; Hoelter 0.05 = 358, and Hoelter 0.01 = 459. The outcomes imply that the model adequately represents the sample data. The internal consistency, measured with Cronbach's coefficient alpha, was equal to 0.89.

The results of the correlation show that the overall result of downstream reciprocity positively correlates, although relatively poorly, with orientation ($r = 0.188^{**}$), sense ($r = 0.210^{**}$), and the religious meaning system ($r = 0.296^{**}$). The outcomes allow the conclusion to be drawn that people who show more intense orientation, sense, and religious meaning system believe more strongly in downstream indirect reciprocity. Only the dimension of religious sense predicted positive DIR ($\beta = 0.210$; $t = 3.686$; $p < 0.001$) with $F(295,1) = 13.587$; $p < 0.001$. With respect to

outliers, none of 296 cases was identified as a possible multivariate outlier. In fact, the lowest value of probability was 0.03834. Cook's value (between 0.000 and 0.088) was under the point at which the researcher should be concerned.

Discussion

This study aimed mainly to examine the internal consistency and nomological validity of PoDIRS-6. The outcomes showed that the new measure has a robust CFA level and meets the criteria for a good fit. Moreover, religiosity correlates positively with PoDIRS-6. It may suggest that people characterized by a religious meaning system are strongly motivated, which enables them to believe in positive downstream reciprocal actions. The outcome is reminiscent of other studies (Dai et al. 2018), implying that the behavior of downstream reciprocity may itself be affected by such issues as religion. For example, in one of the experiments, people (the proposers) who were reminded of the presence of God and knew about the level of religiosity of the anonymous responder, forwarded more money to more religious respondents than to their less religious counterparts (Norenzayan and Shariff 2008). Finally, the dimension of religious sense was a predictor of positive DIR. Since one of the central functions of religion is to help people meet their longing for meaning (Galek et al. 2015) and communicate the commitments that serve as psychological mechanisms for establishing indirect reciprocity (Irons 1996), it is comprehensible that a religious interpretation of life in terms of purpose may influence the beliefs typical of positive DIR. The dimension of religious orientation, which refers to the understanding of ourselves and relationships with other people and the world, was not a predictor of DIR. Such an outcome may suggest that a large part of positive DIR depends on religiosity as a motivator to observe norms. In fact, Bennett and Einolf (2017) have noticed that all main religions play a particularly important role in promoting prosocial behavior toward strangers through a belief in a god or gods who reward(s) kindness and punishes selfishness. Nevertheless, a limitation of this study consists in its correlational design which does not imply causation, as well as in the lack of religious characteristics such as religious affiliation or prayer frequency. The variables that apply religious experience could reflect differences in positive DIR attitudes, beliefs, and behaviors.

Table 7 Basic descriptive statistics of measured variables; Pearson's correlation of the intensity of positive downstream indirect reciprocity with orientation scale, sense scale, and religious meaning system ($N = 296$)

	Mean	SD	Skewness	Kurtosis	r DIR
RELIGIOUS ORIENTATION SCALE	2.76	1.34	0.59	-0.56	0.188**
RELIGIOUS SENSE SCALE	3.77	1.46	0.01	-0.92	0.210**
RELIGIOUS MEANING SYSTEM (GENERAL SCORE)	3.26	1.36	0.33	-0.83	0.296**
POSITIVE DOWNSTREAM INDIRECT RECIPROCITY	5.30	1.09	-0.82	1.24	1

** $p < 0.01$

Study 4

Participants

The study comprised 175 young people (125 women—71.4%) aged between 15 and 19. The average age was 17 ($M = 17.09$; $SD = 0.88$). All students below 18 had a parental consent to participate in the study and were reassured that their participation would be confidential. The participants were recruited via school and university visits.

Procedure and Data Analysis

In order to verify a normal data distribution, skewness and kurtosis were calculated. In the next step, a CFA was performed in order to test the original factor structure of PoDIRS-6. The rationale behind this decision was the same as in Study 3.

Measures

The accuracy of PoDIRS-6 was tested with the use of the Scale of Religious Attitude Intensity (developed by Prężyna, and adapted by Śliwak and Bartczuk 2011) which measures the intensity of the individual’s approach towards the religious attitude object (God and, speaking more precisely, the whole supernatural world). The scale consists of 20 statements, 10 of which must be reversed. The Cronbach α value in this study was high, amounting to 0.95.

Results

Skewness and kurtosis were within acceptable limits of ± 2 (Table 8).

CFA estimations with no missing data confirmed a single-factor solution of PoDIRS-6 with fairly strong loadings of .76, .81, .57, .78, .68, and .81, respectively, and a good model fit with the following goodness of fit indicators: $\chi^2(9) = 15.56$, $p = .076$; $CMIN/DF = 1.730$; $GFI = 0.972$; $TLI = 0.977$; $CFI = 0.986$; $RMSEA = 0.065$ (LO 90 = 0.000; HI 90 = 0.117); $PCLOSE = 0.284$; $Hoelter 0.05 = 190$, and $Hoelter 0.01 = 243$. The outcomes imply that the model adequately represents the sample data. Although RMSEA is higher than .05, it still denotes an acceptable model fit, being less than

0.08. PoDIRS-6 showed satisfactory reliability, as measured by a Cronbach’s alpha coefficient equal to 0.875.

Due to the distribution being close to normal, Pearson’s r correlation was applied. The results of the analysis show that the overall result of positive downstream reciprocity correlates moderately with the intensity of religious attitude ($r = 0.394^{**}$). The positive association between both variables allows the conclusion to be drawn that people who believe more strongly in positive DIR show more intense religious attitudes. With respect to outliers, none of the 175 cases was identified as a possible multivariate outlier. In fact, the lowest value of probability was 0.00760. Cook’s value (between 0.000 and 0.149) was under the point at which the researcher should be concerned (less than 1).

Discussion

Study 4 had a twofold aim. Just as in Study 3, its purpose was to assess the internal consistency and nomological validity of PoDIRS-6. First, the results indicated that the scale meets the criteria for appropriate fit. Second, intensity of religious attitude was positively linked to positive DIR. One of the limitations of the present study regards the difficulty in generalizing the outcomes as a scale that measures religious attitude intensity, as it represents a local measurement tool, limited to the Polish religious and cultural context. Nevertheless, both the Scale of Religious Meaning System by Krok (2011, 2014), and the Scale of Religious Attitude Intensity (Śliwak and Bartczuk 2011) highly correlate with other questionnaires well-known and used around the world, such as the Centrality of Religiosity Scale (Huber and Huber 2012), and the Post-Critical Belief Scale (Fontaine et al. 2003).

Study 5

Participants

The study comprised 173 people (145 women—84%). The participants were aged between 14 and 61 ($M = 28.19$; $SD = 9.74$). The study applied PoDIRS-6, and the Moral Foundations Theory (MFT) developed by Haidt and Graham (2007). Respondents under age took part in the study after obtaining their carer’s consent.

Table 8 Basic descriptive statistics of measured variables; Pearson’s correlation of the intensity of positive downstream indirect reciprocity with intensity of religious attitude ($N = 175$)

	M	SD	Skewness	Kurtosis	r DIR
INTENSITY OF RELIGIOUS ATTITUDE	4.43	1.28	−0.53	−0.08	0.394**
POSITIVE DOWNSTREAM INDIRECT RECIPROCITY	4.87	1.16	−0.54	0.43	1

Measures

The Moral Foundations Questionnaire (MFQ) consists of 30 items (Jarmakowski-Kostrzanowski and Jarmakowska-Kostrzanowska 2016) that assess the degree to which an individual endorses each of the five types of moral concerns: Care/Harm (the appraisal of compassion and kindness, and the depreciation of meanness and imposing suffering); Fairness/Cheating (the endorsement of impartiality, egalitarianism, and universal rights, and the derogation of dishonesty and trickery); In-group Loyalty/Betrayal (the valuation of patriotism, sacrifice, special treatment for one's own in-group, and the disdain of betrayal, treason, cowardice, and lack of help, particularly in times of conflict); Authority/Subversion (the respect, awe, and admiration towards legitimate authorities, and the depreciation of disobedience or uppityness); Purity/Degradation (the awe of temperance, chastity, piety, and cleanliness, and the disapproval of dirtiness, lust, intemperance, and impurity) (Haidt and Graham 2007). Participants are asked to rate on a 6-point Likert scale (from 1—"not at all relevant/strongly disagree," to 6—"extremely relevant/strongly agree") and to answer how pertinent each of the statements is to them. The scores determine the participant's endorsement of each moral concern. The reliability result of the whole scale was 0.82.

Results

Pearson's correlation was applied as the skewness and kurtosis were within acceptable limits of ± 2 (Table 9).

As in the previous studies, the structure of PoDIRS-6 was proved through CFA. The loadings of the indicators were analysed. The analysis confirmed a single-factor solution with strong loadings of .73, .81, .72, .87, .82, and .83. Exemplification of goodness of fit shows: $\chi^2(9) = 15.44$, $p = 0.079$; CMIN/DF = 1.716; RMR = 0.036; GFI = 0.979; AGFI = 0.951; NFI = 0.983; RFI = 0.971; IFI = 0.993; TLI = 0.988; CFI = 0.993; RMSEA = 0.055 (LO 90 = 0.000; HI 90 = 0.101). Although the confidence intervals are slightly outside the recommended range, as was the case in study 4, all other indices confirm the good fit of the model. The test of the closeness of fit

(PCLOSE) shows that the model fits the data well (0.376). The Hoelter critical number 0.05 (258) and the Hoelter 0.01 (330) imply that the model adequately represents the sample data.

The Positive Downstream Indirect Reciprocity Scale showed satisfactory reliability, measured by a Cronbach alpha coefficient equal to 0.919. An analysis of the scale structure was repeated in the form of an exploratory factor analysis (without imposing the number of factors) with the principal component method with varimax rotation and Kaiser-Meyer-Olkin standardization. In this case, similarly to previous studies, the statistics indicated a one-factor structure. The KMO was 0.911, and Bartlett's test of sphericity was $\chi^2 = 889.185$, $df = 15$, $p < 0.001$, indicating the adequacy of the sample selection. The total explained variance was almost 70% (69.747).

In accordance with H4, the results show that positive DIR is positively correlated with care, fairness, in-group loyalty, purity, and slightly with authority, allowing it to be assumed that stronger beliefs in downstream indirect reciprocity are associated with compassion and kindness, impartiality and honesty, admiration toward legitimate authorities, valuation of sacrifice, and temperance (Table 8). Such outcomes confirm H4.

Moreover, the results of a linear regression analysis showed that two of five moral concerns were predictors of positive DIR beliefs: Care/Harm ($\beta = 0.217$; $p = 0.001$), and Loyalty/Betrayal ($\beta = 0.168$ $p = 0.012$), with $F(172,2) = 13.101$; $p < 0.001$. With respect to outliers, none of the 296 cases was identified as a possible multivariate outlier. In fact, the lowest value of probability was 0.00106. Cook's value (between 0.000 and 0.091) was under the point at which the researcher should be concerned.

Discussion

Study 5 had the aim to evaluate the internal consistency and nomological validity of PoDIRS-6, and to assess the predictive function of the five types of moral concerns within the context of positive DIR. First, the outcomes showed that the scale meets the criteria of an appropriate fit. Second, the five types of moral concerns correlated positively with positive DIR. Third, two of them relevantly predicted DIR with different explanatory power. Contrary to what was expected, Fairness/Cheating, which

Table 9 Basic descriptive statistics of the measured variables; Pearson's correlation of the intensity of positive downstream indirect reciprocity with moral concerns ($N = 173$)

	M	ME	SD	SKEWNESS	KURTOSIS	r DIR
CARE/HARM	4.96	5.00	0.54	-0.74	0.74	0.270**
FAIRNESS/CHEATING	4.65	4.66	0.55	-0.45	0.12	0.223**
INGROUP LOYALTY/BETRAYAL	3.77	3.83	0.63	-0.24	0.39	0.332**
AUTHORITY/SUBVERSION	3.68	3.66	0.66	-0.24	-0.52	0.241**
PURITY/DEGRADATION	4.01	4.00	0.72	-0.15	-0.24	0.395**
DOWNSTREAM INDIRECT RECIPROCITY	5.06	5.16	1.10	-0.81	0.87	1

includes the endorsement of universal rights and the derogation of dishonesty, did not become a predictor of DIR. The absence of Fairness/Cheating as a predictor of DIR may be due to the fact that fairness-based giving depends on the “fairness criteria” of the potentially reciprocating agents. In fact, Baker (2012) reported that some people always apply “fairness criteria,” others only in extreme cases, and yet others, never. Moreover, although impartiality is important to downstream reciprocity, it seems that it is one of several factors, and not necessary the strongest one, that occur within the context of downstream reciprocity. Interestingly, both Care/Harm, which expresses the appraisal of compassion and the depreciation of meanness, and Loyalty/Betrayal, which relates to the valuation of sacrifice and the disdain of a lack of help, resulted as two predictors of positive DIR. Moral concern for Care shares some characteristics with gratitude, which was found to be one of the strongest predictors of positive DIR across Studies 2–4. Indeed, both caring and grateful dispositions are based on empathy and the appraisal of kindness. There is considerable evidence (Buck 2011) that caring leads to unselfish tendencies to help and benefit others. Likewise, Loyalty/Betrayal as a predictor of DIR sheds new light on positive downstream reciprocity. Loyalty is a moral dimension that consists in integrity despite its costs, and connects people to others (Nesse 2001). Since DIR, in its positive dimension, refers to a belief that human actions, when positive and morally good, can be repaid by a third person, it is plausible to think that the rewarding persons bear the cost of their downstream reciprocity as they are only observers of someone else’s beneficial behavior and not its addressees.

General Discussion

The analyses of positive downstream indirect reciprocity within a psychological framework confirm that little is theoretically known about this construct and its mechanisms. To our knowledge, the present report is the first study presenting a tool to measure belief in positive downstream indirect reciprocity (PoDIRS-6) among adolescents and adults. It is also the first research investigating positive downstream reciprocity in relationship to gratitude, life satisfaction, religiosity, and moral concerns. In the context of this novelty, we would like to address our results from both theoretical and empirical perspectives.

With respect to the theoretical perspective, PoDIRS-6 contains one type of belief which shows that a person who helped somebody in the past is more likely to receive support from other people in the future. The quite detailed analysis of the literature regarding downstream reciprocity presented in the introductory part of the paper showed that this construct is a complex one, and it may be viewed within the context of many different systems of assessing individuals. In the present research, developing a new measurement method, we performed a certain kind of simplification that could result in cutting out some important and subtle facets of the downstream reciprocity

phenomenon, i.e.: reputation or cooperation, the behavioral aspect of DIR, and negative beliefs about downstream reciprocity. It is this last aspect that requires special explanation. When we started working on creating a questionnaire for measuring downstream indirect reciprocity, we intended to include both positive and negative aspects of downstream reciprocity. Both the introductory part of the article and the proposed items within the initial pool of twenty-eight statements contained positive and negative beliefs about what can happen in the future when people help or harm somebody else. However, the psychometric process of building the questionnaire proved, at various stages, that the most appropriate choice would be a 6-item scale, containing items which suggest the positive nature of beliefs that good deeds done to someone return in the future from other people. Moreover, the content of the items chosen in the final analysis (PoDIRS-6) implies that the scale is not supposed to be used to measure action tendencies, although in the original intention, such items were taken into account. For this reason, knowing that downstream reciprocity can also reflect a negative belief about the return of bad deeds in the future, we strongly recommend that an equivalent NeDIRS-6 be developed: a negative version of PoDIRS-6 which includes action-focused items, as well. It is important to notice that the construction of PoDIRS-6 alone does not constitute a meaningful theoretical contribution. Another important aspect of this research is the relevance that two of five moral concerns predicted positive DIR.

Moreover, the correlational results show that the majority of the variables included in the analyses are associated with positive DIR to a moderate level. Such an outcome allows us to assume that the positive DIR concept is theoretically distinct and independent from gratitude, life satisfaction, religiosity, and moral concerns, although it shares some conceptual nuances. This empirical outcome confirms the introductory insights about the conceptual differences between positive DIR and the abovementioned variables.

With regard to the empirical aspects of the current research, all of the studies performed reported a very good internal consistency of the scale, confirming its reliability. However, we acknowledge several limitations of the study. First, we used convenience samples across four of the studies. The lack of heterogeneously recruited groups poses generalizability questions and does not allow the results from the samples to be applied to the general population. Second, we collected data under a cross-sectional study design. Therefore, a longitudinal panel design would permit deeper examination of the processes, and provide more opportunities to identify and explain possible changes in understanding the development of positive DIR beliefs. Moreover, an experimental design would be required to draw causal inference. Third, an important limitation concerned a discrepancy between the CFA outcomes and the EFA results, requiring modification of the original model of the scale. In this case, we followed Anderson and Gerbing’s (1988) recommendations to perform a series of

respecifications based on a combination of statistical and theoretical considerations. The authors underline that when items are deleted from the measure, the new version of the scale should be subjected to another independent sample. For this reason, replications of PoDIRS-6 were performed in Studies 3–5. In all of the studies, the 6-item version brought about a one-factor solution with good fit indices, and the internal consistency was excellent. Therefore, it appears that the scale is a good method for assessment of beliefs that human actions can be rewarded and compensated in the future by a third person or fate when they are positive and morally good. Fourth, we could not provide a more robust demonstration of construct validity as the association between PoDIRS-6 and the gratitude, religiosity, life satisfaction, and moral concerns measures ranged to 0.39. Although the pattern of the correlations was in line with our hypotheses, they were weak and moderate. These outcomes allow us to consider other questionnaires measuring altruistic cooperation and/or punishment, retributive justice, religious prosociality, and forgiveness to verify which of them is theoretically closer to the beliefs illustrating DIR. Intuitively, we propose the concept of belief in a just world as the potentially strongest correlate of positive downstream reciprocity. This is because people confronted by injustice may react differently to observed unfairness (Dalbert 2001), blaming or justifying the wrongdoers. In future studies, it would be important to apply a questionnaire measuring belief in a just world to verify if DIR may be a component of a just-world belief, or may be driven by it as was reported by Edlund et al. (2007).

The reported limitations call for some potential methodological improvements. Thus, future research should consider

larger and more diversified groups of participants, the longitudinal and experimental approach in scale development (Morgado et al. 2017), and various other measures allowing greater comprehension of the dimension of DIR. It would allow generalization of the results and both theoretically and empirically broaden a nomological network of convergent and discriminant constructs of DIR.

Compliance with Ethical Standards

Conflicts of Interest/Competing Interests The authors declare that they have no conflict of interest.

Appendix 1. Positive Downstream Indirect Reciprocity Scale (PoDIRS-6)

Please, carefully read all statements included in the boxes and use the following scale to assess to what extent you agree or disagree with the given statement. There are no good or bad answers. It is important that they are true for you. Mark the appropriate answer with a cross at each statement:

- 1 I strongly disagree
- 2 I disagree
- 3 I rather disagree
- 4 It is indifferent to me
- 5 I rather agree
- 6 I agree
- 7 I strongly agree

No.	Statement	I strongly disagree	I disagree	I rather disagree	It is indifferent to me	I rather agree	I agree	I strongly agree
1 (2)	In helping others, we help ourselves, for whatever good we give out completes the circle and comes back to us.	1	2	3	4	5	6	7
2 (9)	Life is like an echo, we get back what we give.	1	2	3	4	5	6	7
3 (12)	When we help somebody, somebody else will help us.	1	2	3	4	5	6	7
4 (15)	It is worth being good towards others because it will come back to us sooner or later.	1	2	3	4	5	6	7
5 (18)	Good done to somebody else comes back all of a sudden, sometimes even stronger.	1	2	3	4	5	6	7
6 (19)	Good done to others always comes back.	1	2	3	4	5	6	7

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alexander, R. D. (1986). Ostracism and indirect reciprocity: The reproductive significance of humor. *Ethology and Sociobiology*, 7, 253–270. [https://doi.org/10.1016/0162-3095\(86\)90052-X](https://doi.org/10.1016/0162-3095(86)90052-X).
- Alexander, R. D. (1987). *The biology of moral systems*. Hawthorne: Aldine de Gruyter.
- Algoe, S. B., Haidt, J., & Gabie, S. L. (2008). Beyond reciprocity: Gratitude and relationships in everyday life. *Emotion*, 8, 425–429. <https://doi.org/10.1037/1528-3542.8.3.425>.
- Altmann, T., & Roth, M. (2018). The self-esteem stability scale (SESS) for cross-sectional direct assessment of self-esteem stability. *Frontiers in Psychology*, 9, 91. <https://doi.org/10.3389/fpsyg.2018.00091>.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>.
- Baker, W. (2012). A dual model of reciprocity in organizations: Moral sentiments and reputation. In K. S. Cameron & G. M. Spreitzer (Eds.), *The Oxford handbook of positive organizational scholarship* (pp. 412–422). New York: Oxford University Press.
- Balconi, M., Fronda, G., & Vanutelli, M. E. (2020). A gift for gratitude and cooperative behaviour: Brain and cognitive effects. *Social Cognitive and Affective Neuroscience*, 28, 1317–1327. <https://doi.org/10.1093/scan/nsaa003>.
- Bennett, M. R., & Einolf, C. F. (2017). Religion, altruism, and helping strangers: A multilevel analysis of 126 countries. *Journal for the Scientific Study of Religion*, 56, 323–341. <https://doi.org/10.1111/jssr.12328>.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>.
- Bentler, P. M., & Bonnett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588–606. <https://doi.org/10.1037/0033-2909.88.3.588>.
- Blogowska, J., & Saroglou, V. (2011). Religious fundamentalism and limited prosociality as a function of the target. *Journal for the Scientific Study of Religion*, 50, 44–60. <https://doi.org/10.1111/j.1468-5906.2010.01551.x>.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.000149>.
- Boser, U. (2014). *The leap: The science of trust and why it matters*. New York: New Harvest Houghton Mifflin Harcourt.
- Bowen, N. K., & Guo, S. (2012). *Structural equation modelling*. New York: Oxford University Press.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21, 230–258. <https://doi.org/10.1177/0049124192021002005>.
- Buck, R. (2011). Communicative genes in the evolution of empathy and altruism. *Behavior Genetics*, 41, 876–888. <https://doi.org/10.1007/s10519-011-9456-3>.
- Burger, J. M., Horita, M., Kinoshita, L., Roberts, K., & Vera, C. (1997). Effects of time on the norm of reciprocity. *Basic and Applied Social Psychology*, 19, 91–100. <https://doi.org/10.1207/15324839751037165>.
- Bushman, B. J., Ridge, R. D., Das, E., Key, C. W., & Busath, G. L. (2007). When god sanctions killing: Effect of spiritual violence on aggression. *Psychological Science*, 18, 204–207. <https://doi.org/10.1111/j.1467-9280.2007.01873.x>.
- Byrne, B. M. (2008). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. New York: Taylor & Francis.
- Cangur, S., & Erkan, I. (2015). Comparison of model fit indices used in structural equation modelling under multivariate normality. *Journal of Modern Applied Statistical Methods*, 14, 152–167. <https://doi.org/10.22237/jmasm/1430453580>.
- Caprara, G. V., & Steca, P. (2005). Self-efficacy beliefs as determinants of prosocial behaviour conducive to life satisfaction across ages. *Journal of Social and Clinical Psychology*, 24, 191–217. <https://doi.org/10.1521/jscp.24.2.191.62271>.
- Cate, R. M., Lloyd, S. A., & Henton, J. M. (1982). Fairness and reward level as predictors of relationship satisfaction. *Social Psychology Quarterly*, 45, 177–181. www.jstor.org/stable/3033651.
- Chang, Y. P., Lin, Y. C., & Chen, L. H. (2012). Pay it forward: Gratitude in social networks. *Journal of Happiness Studies*, 13, 761–781. <https://doi.org/10.1007/s10902-011-9289-z>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Cohn, E. S., & Modecki, K. L. (2007). Gender differences in predicting delinquent behaviour: Do individual differences matter? *Social Behaviors and Personality*, 35, 359–374. <https://doi.org/10.2224/sbp.2007.35.3.359>.
- Cooley, S., Elenbaas, L., & Killen, M. (2012). Moral judgments and emotions: Adolescents' evaluations in intergroup social exclusion contexts. *New Directions for Youth Development*, 136, 41–49. <https://doi.org/10.1002/yd.20037>.
- Correia, I., Toscano Batista, M., & Lima, M. L. (2009). Does the belief in a just world bring happiness? Causal relationships among belief in a just world, life satisfaction and mood. *Australian Journal of Psychology*, 61, 220–227. <https://doi.org/10.1080/00049530802579515>.
- Cox, J., & Deck, C. (2006). When are women more generous than men? *Economic Inquiry*, 44, 587–598. <https://doi.org/10.2139/ssrn.895709>.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. <https://doi.org/10.1037/h0040957>.
- Dai, J., Zhang, Yiqiang, and Shi, V. (2018). An economic model of human cooperation based on indirect reciprocity and its implication on environmental protection. *International Journal of Environmental Research and Public Health*, 15, pii: E1303. <https://doi.org/10.3390/ijerph15071303>.
- Dalbert, C. (2001). *The justice motive as a personal resource: Dealing with challenges and critical life events*. New York: Plenum Press.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- de Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305, 1254–1258. <https://doi.org/10.1126/science.1100735>.

- DeSteno, D., Duong, F., Lim, D., & Kates, S. (2019). The grateful don't cheat: Gratitude as a fount of virtue. *Psychological Science*, *30*, 979–988. <https://doi.org/10.1177/0956797619848351>.
- Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL*. London: Sage.
- Diener, E., Emmons, R., Larsen, J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75. https://doi.org/10.1207/s15327752jpa4901_13.
- Dittrich, M. (2015). Gender differences in trust and reciprocity: Evidence from a large-scale experiment with heterogeneous subjects. *Applied Economics*, *47*, 1–14. <https://doi.org/10.1080/00036846.2015.1019036>
- Edlund, J. E., Sagarin, B. J., & Johnson, B. S. (2007). Reciprocity and the belief in a just world. *Personality and Individual Differences*, *43*, 589–596.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785–791. <https://doi.org/10.1038/nature02043>.
- Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, *99*, 689–723. <https://doi.org/10.1037/0033-295X.99.4.689>
- Flannelly, K. J., & Galek, K. (2006). Discipline and sex differences in religiosity and spirituality among health care professionals. *Psychological Reports*, *99*, 803–804. <https://doi.org/10.2466/PR.99.3.803-804>.
- Fontaine, J., Duries, B., Luyten, P., & Hutsebaut, D. (2003). The internal structure of the post-critical belief scale. *Personality and Individual Differences*, *35*, 501–518. [https://doi.org/10.1016/S0191-8869\(02\)00213-1](https://doi.org/10.1016/S0191-8869(02)00213-1).
- Francis, L. J. (1992). Religion, neuroticism, and psychoticism. In J. F. Schumaker (Ed.), *Religion and mental health* (pp. 149–160). New York: Oxford University Press.
- Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T., & Fralish, J. S. (1995). Parallel analysis: A method for determining significant principal components. *Journal of Vegetation Science*, *6*, 99–106. <https://doi.org/10.2307/3236261>.
- Galek, K., Flannelly, K. J., Ellison, C. G., Siltan, N. R., & Jankowski, K. R. B. (2015). Religion, meaning and purpose, and mental health. *Psychology of Religion and Spirituality*, *7*, 1–12. <https://doi.org/10.1037/a0037887>.
- George, J. M. (1991). State or trait: Effects of positive mood on Prosocial behaviors at work. *Journal of Applied Social Psychology*, *76*, 299–307. <https://doi.org/10.1037/0021-9010.76.2.299>.
- George, D., & Mallery, P. (2014). *IBM SPSS statistics step by step: A simple guide and reference*. Boston: Allyn and Bacon.
- Gervais, W. M., & Norenzayan, A. (2012). Like a camera in the sky? Thinking about god increases public self-awareness and socially desirable responding. *Journal of Experimental Social Psychology*, *48*, 298–302. <https://doi.org/10.1016/j.jesp.2011.09.006>.
- Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, *55*, 377–393. <https://doi.org/10.1177/0013164495055003002>.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*, 98–116. <https://doi.org/10.1007/s11211-007-0034-z>.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modelling (PLS-SEM)*. Thousand Oaks: Sage.
- Harper, D. J., Wagstaff, G. F., Newton, T., & Harrison, K. R. (1990). Lay causal perceptions of third world poverty and the just world theory. *Social Behaviors and Personality*, *18*, 235–238. <https://doi.org/10.1114/sbp.1990.18.2.235>.
- Heintz, S., Kramm, C., & Ruch, W. (2017). A meta-analysis of gender differences in character strengths and age, nation, and measure as moderators. *The Journal of Positive Psychology*, *14*, 103–112. <https://doi.org/10.1080/17439760.2017.1414297>.
- Hemphill, J. F. (2003). Interpreting the magnitude of correlation coefficients. *The American Psychologist*, *58*, 78–79. <https://doi.org/10.1037/0003-066X.58.1.78>.
- Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods & Research*, *11*, 325–344. <https://doi.org/10.1177/0049124183011003003>.
- Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and just what they do. *PNAS*, *112*, 1727–1732. <https://doi.org/10.1073/pnas.1417904112>.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*, 53–60. <https://doi.org/10.21427/D7CF7R>.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modelling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks: Sage.
- Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. <https://doi.org/10.1080/10705519909540118>.
- Huber, S., & Huber, O. W. (2012). The centrality of religiosity scale (CRS). *Religions*, *3*, 710–724. <https://doi.org/10.3390/rel3030710>.
- Irons, W. (1996). Morality, religion, and human evolution. In W. M. Richardson & W. J. Wildman (Eds.), *Religion & Science: History, method, dialogue* (pp. 375–399). New York: Routledge.
- Janoff-Bulman, R., & Carnes, N. C. (2013). Surveying the moral landscape: Moral motives and group-based moralities. *Personality and Social Psychology Review*, *17*, 219–236. <https://doi.org/10.1177/108886831348274>.
- Jarmakowski-Kostrzanowski, T., & Jarmakowska-Kostrzanowska, L. (2016). Polska adaptacja *Kwestionariusza kodów moralnych* (MFQ-PL). *Psychologia Społeczna*, *11*, 489–508. <https://doi.org/10.7366/1896180020163908>.
- Jones, S., Raykos, B. C., McEvoy, P. M., Ieraci, J., Fursland, A., Byrne, S., & Waller, G. (2018). The development and validation of a measure of eating disorder-specific interpersonal problems: The interpersonal relationships in eating disorders (IR-ED) scale. *Psychological Assessment*, *6*, 389–403. <https://doi.org/10.1037/pas0000666>.
- Jorgensen, T. D., Kite, B. A., Chen, P. Y., & Short, S. D. (2018). Permutation randomization methods for testing measurement equivalence and detecting differential item functioning in multiple-group confirmatory factor analysis. *Psychological Methods*, *23*, 708–728. <https://doi.org/10.1037/met0000152>.
- Juczyński, Z. (2009). *Narzędzia pomiaru w promocji i psychologii Zdrowia*. Warszawa: Pracownia Testów Psychologicznych.
- Kaczmarek, Ł. D., Kashdan, T. B., Drajkowski, D., Enko, J., Kosakowski, M., Szäefer, A., & Bujacz, A. (2015). Why do people prefer gratitude journaling over gratitude letters? The influence of individual differences in motivation and personality on web-based interventions. *Personality and Individual Differences*, *75*, 1–6. <https://doi.org/10.1016/j.paid.2014.11.004>.
- Karadağ, S. Ç., & Akgun, S. (2016). A cross-cultural comparison of Turkish and Kyrgyz students' belief in a just world. *Journal of*

- Human Sciences*, 13, 4229–4236. <https://doi.org/10.14687/jhs.v13i3.3987>.
- Karbowski, A. (2011). O kilku modelach samolubnego karania w ekonomii behawioralnej. *International Journal of Management and Economics*, 29, 236–249. <https://doi.org/10.0017/1211555>.
- Kashdan, T. B., Mishra, A., Breen, W. E., & Froh, J. J. (2009). Gender differences in gratitude: Examining appraisals, narratives, the willingness to express emotions, and changes in psychological needs. *Journal of Personality*, 77, 691–730. <https://doi.org/10.1111/j.1467-6494.2009.00562.x>.
- Kim I., & Kim, C. S. (2003). Patterns of family support and the quality of life of the elderly. *Social Indicators Research*, 62, 437–454. <https://doi.org/10.1023/A:1022617822399>
- Kopalle, P. K., Lehmann, D. R., & Farley, J. U. (2010). Consumer expectations and culture: Effect of belief in karma in India. *Journal of Consumer Research*, 37, 251–263. <https://doi.org/10.1086/651939>.
- Kossakowska, M., & Kwiatek, P. (2014). Polska adaptacja kwestionariusza do badania wdzięczności GQ-6. *Przegląd Psychologiczny*, 57, 503–514.
- Krok, D. (2011). Skala Religijnego Systemu Znaczeń (SRSZ). In M. Jarosz (Ed.), *Psychologiczny pomiar religijności* (pp. 153–168). Lublin: Towarzystwo Naukowe KUL.
- Krok, D. (2014). Religijny system znaczeń i religijne radzenie sobie ze stresem a eudajmonistyczny dobrostan psychiczny. *Roczniki Psychologiczne*, 17, 647–664. E-ISSN: 2451–4306.
- Kulow, K., & Kramer, T. (2016). It pursuit of good karma: When charitable appeals to do right go wrong. *Journal of Consumer Research*, 43, 334–353. <https://doi.org/10.1093/jcr/ucw018>.
- Kyeong, S., Kim, J., Kim, D. J., Kim, H. E., & Kim, J. J. (2017). Effects of gratitude mediation on neural network functional connectivity and brain-heart coupling. *Scientific Reports*, 11(7), 5058. <https://doi.org/10.1038/s41598-017-05520-9>.
- Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. New York: Plenum Press.
- Liao, S. M. (2016). *Moral brains: The neuroscience of morality*. New York: Oxford University Press.
- Lucas, T., Zhdanova, L., Wendorf, C. A., & Alexander, S. (2013). Procedural and distributive justice beliefs for self and others: Multilevel associations with life satisfaction and self-rated health. *Journal of Happiness Studies*, 14, 1325–1341. <https://doi.org/10.1007/s10902-012-9387-6>.
- Ma, L. K., Tunney, R. J., & Ferguson, E. (2017). Does gratitude enhance prosociality? A meta-analytic review. *Psychological Bulletin*, 143, 601–635. <https://doi.org/10.1037/bul0000103>.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size of covariance structure modeling. *Psychological Methods*, 1, 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>.
- McAleer, S. (2016). Are gratitude and forgiveness symmetrical? In D. Carr (Ed.), *Perspectives on gratitude: An interdisciplinary approach* (pp. 85–96). New York: Routledge.
- McClelland, N. C. (2010). *Encyclopedia of reincarnation and karma*. Jefferson: McFatland & Company.
- McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? *Psychological Bulletin*, 127, 249–266. <https://doi.org/10.1037/0033-2909.127.2.249>.
- Melas, C. D., Zampetakis, L. A., Dimopoulou, A., & Moustakis, V. (2011). Modeling the acceptance of clinical information systems among hospital medical staff: An extended TAM model. *Journal of Biomedical Informatics*, 44, 553–564. <https://doi.org/10.1016/j.jbi.2011.01.009>.
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3, 111–121. <https://doi.org/10.21500/20112084.857>.
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexo e Critica*, 30, 3. <https://doi.org/10.1186/s41155-016-0057-1>.
- Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation Modelling. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 488–508). Thousand Oaks: Sage.
- Muthén, B., & Kaplan, D. (1985). A comparison of methodologies for the factor analysis of non-Normal Likert variables. *The British Journal of Mathematical and Statistical Psychology*, 38, 171–189. <https://doi.org/10.1111/j.2044-8317.1985.tb00832.x>.
- Nesse, R. M. (2001). Natural selection and the capacity for subjective commitment. In R. M. Nesse (Ed.), *Evolution and the capacity for commitment* (pp. 1–44). New York: Russell Sage Foundation.
- Ng, J. W. X., Tong, E. M. W., Sim, D. L. Y., Loy, X., & Giesbrecht. (2017). Gratitude facilitates private conformity: A test of the social alignment hypothesis. *Emotion*, 17, 379–387. <https://doi.org/10.1037/emo0000249>.
- Niesta Kayser, D., Greitemeyer, T., Fischer, P., & Frey, D. (2011). Why mood affects help giving, but not moral courage: Comparing two types of prosocial behaviour. *European Journal of Social Psychology*, 40, 1136–1157. <https://doi.org/10.1002/ejsp.717>.
- Nordin, A. (2015). Indirect reciprocity and reputation management in religious morality relating to concepts of supernatural agents. *Journal for the Cognitive Science of Religion*, 3, 125–153. <https://doi.org/10.1558/jcsr.27256>.
- Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science*, 322, 58–62. <https://doi.org/10.1126/science.1158757>.
- Nowak, M. A., & Highfield, R. (2011). *SuperCooperators: Altruism, evolution, and why we need each other to succeed*. New York: Free Press.
- Nowak, M. A., & Roch, S. (2007). Upstream reciprocity and evolution of gratitude. *Proceedings of the Royal Society B: Biological Sciences*, 274, 605–609. <https://doi.org/10.1098/rspb.2006.0125>.
- Nowak, M. A., & Sigmund, K. (1998a). Evolution of indirect reciprocity by image scoring. *Nature*, 393, 573–577. <https://doi.org/10.1038/31225>.
- Nowak, M. A., & Sigmund, K. (1998b). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, 194, 561–574. <https://doi.org/10.1006/jtbi.1998.0775>.
- Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, 437, 1291–1298. <https://doi.org/10.1038/nature04131>.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32, 396–402. <https://doi.org/10.3758/bf03200807>
- O'Connor, W. E., Morrison, T. G., McLeod, L. D., & Anderson, D. (1996). A meta-analytic review of the relationship between gender and belief in a just world. *Journal of Social Behavior and Personality*, 11, 141–148.
- Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. Scotts Valley: CreateSpace Independent Publishing.
- Osgood, J. M. (2017). Is revenge about retributive justice, deterring harm, or both? *Social and Personality Psychology Compass*, 11, e112296. <https://doi.org/10.1111/spc3.12296>.
- Park, G. P. (2014). Factor analysis of the foreign language classroom anxiety scale in Korean learners of English as a foreign language. *Psychological Reports*, 115, 261–275. <https://doi.org/10.2466/28.11.PR0.115c10z2>.
- Peng, C., Malafosse, C., Nelissen, R. M. A., & Zeelenberg, M. (2019). Gratitude, indebtedness, and reciprocity: An extended replication of Bartlett & DeSteno (2006). *Social Influence*, 15, 4–16. <https://doi.org/10.1080/15534510.2019.1710248>.

- Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2002). The personal norm of reciprocity. *European Journal of Personality, 17*, 251–283. <https://doi.org/10.1002/per.474>.
- Peter, J. P. (1981). Construct validity: A review of basic issues and marketing practices. *Journal of Marketing Research, 18*, 133–145.
- Pohling, R., Diessner, R., & Strobel, A. (2017). The role of gratitude and moral elevation in moral identity development. *International Journal of Behavioral Development, 42*, 405–415. <https://doi.org/10.1177/0165025417727874>.
- Polit, D. F., & Tatano Beck, C. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*, 489–497. <https://doi.org/10.1002/nur.20147>.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>.
- Reich, B., & Wang, X. (2015). And justice for all: Revisiting the global belief in a just world scale. *Personality and Individual Differences, 78*, 68–76. <https://doi.org/10.1016/j.paid.2015.01.031>.
- Reich, B. J., Beck, J. T., & Price, J. (2018). Food as ideology: Measurement and validation of Locavorism. *Journal of Consumer Research, 45*, 849–868. <https://doi.org/10.1093/jcr/ucy027>.
- Righi, S., & Takács, K. (2018). Social closure and the evolution of cooperation via indirect reciprocity. *Scientific Reports, 8*, 11149. <https://doi.org/10.1038/s41598-018-29290-0>.
- Rockenbach, B., & Milinski, M. (2006). The efficient interaction of indirect reciprocity and costly punishment. *Nature, 444*, 718–723. <https://doi.org/10.1038/nature05229>.
- Roth, L. M., & Kroll, J. C. (2007). Risky business: Assessing risk preferences explanations for gender differences in religiosity. *American Sociological Review, 72*, 205–220. <https://doi.org/10.1177/000312240707200204>.
- Schinka, J. A., Velicer, W. F., & Weiner, I. R. (2012). *Handbook of psychology: Research methods in psychology*. Hoboken: Wiley.
- Schmitt, T. A. (2011). Current methodological consideration in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29*, 304–321. <https://doi.org/10.1177/0734282911406653>.
- Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*, 323–337. <https://doi.org/10.3200/JOER.99.6.323-338>.
- Shen, F. C., Wang, Y. W., & Swanson, J. L. (2011). Development and initial validation of the internalization of Asian American stereotypes scale. *Cultural Diversity & Ethnic Minority Psychology, 17*, 283–294. <https://doi.org/10.1037/a0024727>.
- Śliwak, J., & Bartczuk, R. (2011). Skala Intensywności Postawy Religijnej. In M. Jarosz (Ed.), *Psychologiczny pomiar religijności* (pp. 153–168). Lublin: Towarzystwo Naukowe KUL.
- Stavrova, O., & Siegers, P. (2014). Religious prosociality and morality across cultures: How social enforcement of religion shapes the effects of personal religiosity on prosocial and moral attitudes and behaviors. *Personality and Social Psychology Bulletin, 40*, 315–333. <https://doi.org/10.1177/014667213510951>.
- Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema, 30*, 149–158. <https://doi.org/10.7334/psicothema2018.33>.
- Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology, 74*, 482–493. <https://doi.org/10.1037/0022-3514.74.2.482>.
- Szcześniak, M. (2018). Indirect reciprocity: The concept and psychological mechanisms. *Roczniki Psychologiczne, 21*, 107–130. <https://doi.org/10.18290/rpsych.2018.21.2-1>.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. New York: Allyn and Bacon.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10. <https://doi.org/10.1007/BF02291170>.
- Tullberg, J. (2004). On indirect reciprocity: The distinction between reciprocity and altruism, and a comment on suicide terrorism. *American Journal of Economics and Sociology, 63*, 1193–1221. <https://doi.org/10.1111/j.1536-7150.2004.00341.x>.
- Tullberg, J. (2012). The golden rule of benevolence versus the silver rule of reciprocity. *Journal of Religion and Business Ethics, 3*. Available from: <http://via.library.depaul.edu/jrbe/vol3/iss1/2>.
- Uchida, S., & Sigmund, K. (2010). The competition of assessment rules for indirect reciprocity. *Journal of Theoretical Biology, 263*, 13–19. <https://doi.org/10.1016/j.jtbi.2009.11.013>.
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *The European Journal of Developmental Psychology, 9*, 486–492. <https://doi.org/10.1080/17405629.2012.686740>.
- Vayness, J., Doung, F., & De Steno, D. (2019). Gratitude increases third-party punishment. *Cognition and Emotion, 8*, 1–8. <https://doi.org/10.1080/02699931.2019.1700100>.
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2005). *Measurement in nursing and health research*. New York: Springer.
- Wenzel, M., & Okimoto, T. G. (2016). Retributive justice. In C. Sabbagh & M. Schmitt (Eds.), *Handbook of social justice theory and research* (pp. 237–256). New York: Springer.
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York: The Guilford Press.
- Williams, L. J., & McGonagle, A. K. (2016). Four research designs and a comprehensive analysis strategy for investigating common method variance with self-report measures using latent variables. *Journal of Business and Psychology, 31*, 339–359. <https://doi.org/10.1007/s10869-015-9422-9>.
- Williams, L. J., & O'Boyle, E. H. (2015). Ideal, nonideal, and no-marker variables: The confirmatory factor analysis (CFA) marker technique works when it matters. *The Journal of Applied Psychology, 100*, 1579–1602. <https://doi.org/10.1037/a0038855>.
- Wood, N. D., Akloubou Gnonhosou, D. C., & Bowling, J. (2015). Combining parallel and exploratory analysis in identifying relationship scales in secondary data. *Marriage & Family Review, 51*, 385–395. <https://doi.org/10.1080/01494929.2015.1059785>.
- Wu, J. B., Hom, P. W., Tetrick, L. E., Shore, L. M., Jia, L., Li, C., & Jiwen Song, L. (2006). The norm of reciprocity: Scale development and validation in the Chinese context. *Management and Organization Review, 2*, 377–402. <https://doi.org/10.1111/j.1740-8784.2006.00047.x>.
- Yaghmaie, F. (2003). Content validity and its estimation. *Journal of Medical Education, 3*, 25–27. <https://doi.org/10.22037/jme.v3i1.870>.
- Yoeli, E., Hoffman, M., Rand, D. G., & Nowak, M. A. (2013). Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of The National Academy of Sciences, 110*, 10424–10429. <https://doi.org/10.1073/pnas.1301210110>.
- Yong, A. G., & Pearce, S. (2013). A beginner's guide for factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*, 79–94. <https://doi.org/10.20982/tqmp.09.2.p079>.