Theoretical and Methodological Challenges in Cross-Cultural Social Psychology Research

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Philosophie an der Ludwig-Maximilians-Universität

vorgelegt von

Zoe Magraw-Mickelson

aus USA

2019

Theoretical and Methodological Challenges in Cross-Cultural Social Psychology Research

Inauguraldissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von
Zoe Magraw-Mickelson

aus USA
2019

Erstgutachter: Prof. Dr. Mario Gollwitzer
Zweitgutachter: Prof. Dr. Dieter Frey
Datum der mündlichen Prüfung: Mai 16, 2019

Table of Contents

Abstract

Social psychology depends on assumptions about the cultural context in which an experiment takes place. Cross-cultural psychology identifies and explains differences not accounted for in these assumptions. However, comparability does not come easy; there are numerous challenges that come with expanding the surrounding context both in the development of theory and execution of methods. This dissertation will review the history of cross-cultural psychology and strive to experimentally rectify some of its shortcomings. Part one will take on the theory behind one of the most researched cross-cultural topics: collectivism, which is traditionally contrasted to individualism. I further differentiate between relational and categorical collectivism, and investigate similarities and differences at a country level as well as at the individual level of self-construal. Part two will address methodological issues in cross-cultural psychology: 1) by looking at the mode in which data is collected; comparing paper/pencil, web/computer-based, and smart-phone, and finding differences across modes which are not explained by country or age alone, and 2) by exploring the fundamental comparability of psychological measures with invariance testing, and by using different methods of testing, such as multiple-group confirmatory factor analysis and alignment, for measurement invariance with the goal to facilitate comparing measures across numerous countries. In conclusion, these experimental explorations both improve and draw attention to the limitations in cross-cultural social psychology.

Zusammenfassung der Dissertation

Die Sozialpsychologie ist zu Annahmen über den kulturellen Kontext gezwungen, in welchem ein Experiment stattfindet. Die kulturvergleichende Psychologie identifiziert und erklärt Unterschiede, die in diesen Annahmen häufig nicht berücksichtigt sind. Vergleichbarkeit ist dabei ein zentrales Thema und kann selten einfach angenommen werden; mit der Vergrößerung des einbezogenen Kontextes entstehen zahlreiche Herausforderungen, sowohl bezüglich der Entwicklung von Theorien, als auch in der Anwendung von Methoden. Diese Dissertation leistet eine Übersicht über die Geschichte der kulturvergleichenden Psychologie und trägt mit einem experimentellen Ansatz dazu bei, einige ihrer Defizite zu korrigieren. Der erste Teil der Arbeit beschäftigt sich mit einem der einschlägigsten kulturvergleichenden Themen: Kollektivismus, welcher traditionell dem Individualismus gegenübergestellt wird. Darüber hinaus unterscheide ich zwischen relationalem und kategorialem Kollektivismus und untersuche Gemeinsamkeiten und Unterschiede auf Länderebene, sowie auf einem individuellen Level der Selbstrepräsentation. Teil zwei adressiert methodologische Herausforderungen der kulturvergleichenden Psychologie: Erstens, indem Modi der Datenerhebung (schriftliche Befragung, Befragung am Computer, Befragung am Smartphone) betrachtet und Unterschiede zwischen ihnen identifiziert werden, welche nicht auf Herkunftsland oder Alter der Teilnehmenden zurückgeführt werden können. Zweitens, indem die fundamentale Vergleichbarkeit von psychologischen Messinstrumenten per Invarianztestung exploriert wird. Hierbei werden verschiedene Methoden der Invarianz-Testung (die „Multiple-group confirmatory factor analysis“ und die „Alignment“-Methode) verglichen, um das Vergleichen von Messwerten über mehrere Länder hinweg zu erleichtern. Schlussendlich machen diese experimentellen Untersuchungen auf bestimmte Limitationen der kulturvergleichenden Psychologie aufmerksam und verringern diese darüber hinaus bedeutend.

Kapitel 1: „Eine Übersicht über das Feld: Was bislang getan wurde und was wir in Zukunft tun müssen“: Was ist kulturvergleichende Sozialpsychologie und wieso ist sie von Bedeutung? Ein kontextueller Vergleich führt nicht nur zu einem besseren Verständnis der Ergebnisse der traditionellen Sozialpsychologie, welche vor allem in westlichen Gesellschaften gefunden wurden, sondern erweitert auch unseren Blick und unser Verständnis über Grenzen hinweg. Gleichwohl ist das Erreichen dieser Ziele leichter gesagt als getan und die Basis auf der wir aufbauen möchten ist nicht so stabil wie wir hoffen würden. Wie in allen Bereichen der Psychologie muss die kulturvergleichende Sozialpsychologie fundamentale Fragen zur Replizierbarkeit ihrer Ergebnisse und ihrer theoretischen Tiefe beantworten. In Kapitel 1 wird zunächst die Geschichte der kulturvergleichenden Sozialpsychologie skizziert, wobei auf die Individualismus- versus Kollektivismus-Dichotomie als Grundpfeiler vieler Theorien der kulturvergleichenden Psychologie fokussiert wird. Hiernach werden die methodologischen Herausforderungen von cross-kulturellen Vergleichen diskutiert, welche die Validierung von Befunden erschweren.

Kapitel 2: „Kulturell bedingte Unterschiede in relationaler und gruppenkollektiver Selbstrepräsentation und ihre Bedeutung für Reaktionen auf beobachtete Ungerechtigkeit“: Mentale Repräsentationen des „Selbst“ bestehen sowohl aus individuellen Aspekten (also inwiefern sich jemand von anderen Menschen unterscheidet) und kollektiven Aspekten (also in welcher Beziehung man zu anderen Menschen steht). Kollektive Aspekte bestehen wiederum aus interpersonellen Beziehungen (dem „relationalen Selbst“) und der Zugehörigkeit zu sozialen Gruppen (dem „gruppenkollektiven Selbst“). Manche Forschende gehen von einer universalen motivationalen Hierarchie in den Selbstrepräsentationen aus (mit dem relationalen Selbst als relevanter als das gruppenkollektive Selbst). Andere Forschung legt nahe, dass die relative Wichtigkeit der verschiedenen Selbstrepräsentationen zwischen Kulturen variiert. In Kapitel 2 teste ich die Hypothese der motivationalen Hierarchie der Selbstrepräsentationen cross-kulturell. Hierfür wurden emotionale Reaktionen (Ärger, Wut

und Racheintentionen) auf die Beobachtung einer Viktimisierung eines kollektiven oder relationalen Gruppenmitgliedes in Deutschland, Japan und den USA untersucht. Im Einklang mit der Vermutung einer konsistenten Hierarchie über Kulturen hinweg fanden wir, dass in allen drei Ländern das relationale Selbst vorrangiger ist als das gruppenkollektive Selbst.

Kapitel 3: „Erweiterung des Selbst bei der Reaktion auf Ungerechtigkeit gegenüber anderen: Kategorisches versus relationales Selbst“: Wann beschäftigt Menschen die Ungerechtigkeit gegenüber anderen am meisten? Aufbauend auf Befunden der Selbstrepräsentations-Literatur stellen wir die Hypothese auf, dass Menschen die Ungerechtigkeit gegenüber anderen am meisten beschäftigt, wenn eine Passung zwischen ihrer Selbstrepräsentation und der Art der Beziehung zum Opfer besteht. Das heißt, dass Beobachter von Ungerechtigkeit am meisten Empörung erleben sollten, wenn bei ihnen eine relationale Selbstrepräsentation gebahnt wird und das Opfer ein relationales Gruppenmitglied ist; oder wenn bei ihnen eine individuelle Selbstrepräsentation gebahnt wird und das Opfer ein Gruppenmitglied ist, mit dem man dieselbe Kategorie teilt. In einer experimentellen Laborstudie mit chinesischen Studierenden (N = 141) verwendeten wir eine neue Methode zur Manipulation der relationalen sowie der kategorischen Gruppen unter Verwendung eines „Minimal Group“ Paradigmas. Darüber hinaus manipulierten wir die Selbstrepräsentationen der Teilnehmenden. Auch wenn wir unter Betrachtung der Manipulation der Selbstrepräsentationen keine Evidenz für unsere primären Hypothesen finden konnten, zeigte sich doch eine indirekte Bestätigung in der gemessenen Selbstrepräsentation der Teilnehmenden. So zeigte sich eine Interaktion zwischen einer interdependenten Selbstrepräsentation der Teilnehmenden und dem Typ des kollektiven Bezugs zum Opfer in der Vorhersage der Empörung bei der Beobachtung der Viktimisierung.

Kapitel 4: „Fragebogenmodi und Datenqualität: Ein Blick auf sorgloses Ausfüllen in drei Modi und drei Kulturen“: Ein Großteil psychologischer Forschung hängt von der Sorgfalt von Versuchsteilnehmenden beim Ausfüllen von Materialien (wie Tests und Fragebögen) ab.

Gleichwohl sind nicht alle Teilnehmenden motiviert das Material sorgfältig zu bearbeiten, wodurch es zu häufig unbeachteten Problemen in der Datenqualität kommt. Unsere Frage lautet: Wie beeinflussen unterschiedliche Modi der Datenerhebung (schriftliche Befragung, Befragung am Computer, Befragung am Smartphone) die Sorgfalt der Teilnehmenden und dadurch die Datenqualität? Ergebnisse früherer Studien legen nahe, dass die unterschiedlichen Modi der Datenerhebung eine vergleichbare Häufigkeit von sorglosen Antworttendenzen mit sich bringen. Da sich die Technologie jedoch weiterentwickelt und Daten in immer heterogeneren Populationen gesammelt werden, muss diese Frage erneut angegangen und darüber hinaus um die Berücksichtigung kultureller Unterschiede erweitert werden. Die vorliegende Forschung untersucht die Effekte der Modi der Datenerhebung auf sorglose Bearbeitung der Materialien in drei Wellen eines messwiederholten Designs. Empfehlungen aus der Literatur folgend, berechneten wir einen Index, der aus acht Indikatoren zusammengesetzt ist, die jeweils Aspekte sorglosen Bearbeitens erfassen. In einer Stichprobe von Erwachsenen im Arbeitsleben in China fanden wir, dass Versuchsteilnehmende signifikant weniger sorgfältig waren wenn sie das Untersuchungsmaterial Online bzw. am Computer bearbeitet haben, im Vergleich zu einer Bearbeitung mit Papier und Stift. In einer studentischen deutschen Stichprobe zeigten sich Teilnehmende wiederum signifikant weniger sorgfältig, wenn sie das Material mit Papier und Stift bearbeitet haben, im Vergleich zu einer Bearbeitung am Smartphone. In einer Stichprobe von chinesisch sprechenden Studierenden fanden wir hingegen keine Unterschiede zwischen den Modi der Datenerhebung. In diesem Kapitel wird diskutiert, wieso unsere Ergebnisse von vergangenen Ergebnissen zum Einfluss von Modi der Datenerhebung abweichen. Darüber hinaus werden Hypothesen über mögliche kulturelle Unterschiede aufgestellt.

Kapitel 5: „Methoden der Messinvarianztestung: Acht Kulturen, acht Messinstrumente“: Messinvarianz (MI) ist unabdingbar für jegliche Vergleiche heterogener Gruppen. Die „Multiple-group confirmatory factor analysis“ (MG-CFA) ist die Standard-

Methode um Messinvarianz zu testen, geht aber mit wohlbekannten Limitationen einher, insbesondere bei einer großen Anzahl von Gruppen, bei denen eine strikte Invarianz im Sinne dieser Methode schwer bis unmöglich zu erreichen ist. Neuere Methoden, insbesondere die „Alignment"-Optimierungsmethode, bringen eine erhöhte Flexibilität und neue Möglichkeiten für Vergleiche zwischen einer hohen Anzahl von Gruppen mit sich. Dieses Kapitel vergleicht MG-CFA mit der Alignment-Methode zur MI-Testung anhand eines Beispiels von acht Ländern, sowie acht verschiedenen Messinstrumenten. Es zeigt sich, dass in Situationen in denen strikte MI mit der MG-CFA problematisch ist, die Alignment-Methode Vergleiche der Gruppen ermöglicht. Darüber hinaus vergleichen wir die inhaltlichen Implikationen der relativen Differenzen zwischen den von den unterschiedlichen Methoden ausgegebenen Mittelwerten. In diesem Vergleich zeigt sich, dass es in den meisten der acht Messinstrumente bei Verwendung der Alignment-Methode weniger signifikante latente Mittelwertsunterschiede gibt als bei Verwendung der MG-CFA. Insgesamt zeigt sich, dass die Alignment-Methode eine vielversprechende Lösung darstellt, um mehr Vertrauen in die Vergleichbarkeit von Messinstrumenten in gruppenvergleichender Forschung zu gewährleisten.

Kapitel 6: Schlussfolgerung: Mithilfe einer großen Vielfalt von Studien ist in der vorliegenden Arbeit eine umfangreiche Betrachtung wichtiger Themen der kulturvergleichenden Sozialpsychologie gelungen. Die Beiträge dieser Arbeit sind vielfältig: 1. Aufzeigen von Ähnlichkeit in der Hierarchie der Selbstrepräsentationen, für welche Theorien der kulturvergleichenden Psychologie Unterschiede annehmen würden. 2. Während die meiste Forschung bisher lediglich Fragebogen verwendet hat, wurde in einer Laborstudie die relationale Gruppenzugehörigkeit manipuliert, wobei die Ergebnisse insbesondere auf die Bedeutung der gemessenen Selbstrepräsentation (im Vergleich zur manipulierten Selbstrepräsentation) hindeuten. 3. Eine Untersuchung des Zusammenspiels von Kulturunterschieden und sorgloser Studienteilnahme. Hierbei zeigten sich Unterschiede

zwischen den Modi der Datenerhebung, welche abhängig von der Nationalität und dem Alter der Teilnehmenden sind. Zuletzt wurden ältere und neuere Methoden verglichen, welche zum Sicherstellen von Messinvarianz genutzt werden; eine Voraussetzung für jedweden Gruppenvergleich. Die damit genommenen Hürden sind wichtig und – obwohl noch mehr zu tun ist – ist jede dieser genommenen Hürden ein bedeutender Schritt auf dem Weg zu einem besseren Verständnis der kulturvergleichenden Sozialpsychologie.

I. Chapter 1: A Review of the Field- What has Been Done and What We Need to Do

Abstract

What is cross-cultural social psychology and why does it matter? Having a contextual comparison not only leads to a better understanding of conclusions of traditional social psychology, which have been explored primarily in Western societies, but also expands our view and comprehension beyond borders. However, achieving these goals is easier said than done, and the base from which we want to build is not as sturdy as we would hope. As in all areas of psychology, cross-cultural social psychology needs to ask fundamental questions about the replicability of our results and the depth of our theory. Chapter 1 will go over a brief history of cross-cultural social psychology, focusing on the common theme of individualism versus collectivism dichotomy as a pillar of theory in cross-cultural psychology. This will be followed by addressing the methodological challenges of cross-cultural comparisons that hinder the verifiability of findings.

What is cross-cultural psychology?

Social psychology investigates how social interactions shape how we think, feel, and act, that is, how our existence as social animals shapes our cognitive reality. Cross-cultural social psychology looks at how cultures further influence these processes. What is culture? Simply put, culture is a shared system, produced and perpetuated by a group. To further our understanding, we look to others in the field who define culture as the following:

> "The collective programming of the mind that distinguishes the members of one group or category of people from another" (Hofstede, 2001, p. 9)

> "… a unique meaning and information system, shared by a group and transmitted across generations, that allows the group to meet basic needs of survival, pursue happiness and well-being, and derive meaning from life." (Matsumoto & Juang, 2008, p. 12)

> "… it [culture] represents a coalescence of discrete behavioral norms and cognitions shared by individuals within some definable population that are distinct from those shared within other populations. These normative beliefs and behaviors provide resources for realizing individual and collective goals, and so are often institutionalized in a variety of formal and informal ways. Moreover, there exist means for transmitting beliefs and behaviors to new members of the cultural population, so that the norms defining a culture may persist over very long periods of time." (Lehman, Chiu, & Schaller, 2004 p. 690)

Cross-cultural social psychology (CCSP) is in the fabric of social psychology itself, the influences of which are defined as culture aligned with what social psychologists seek to

explore. However, CCSP is differentiated by comparing two or more different cultural contexts or groups. Cultural groups are not necessary defined by national boarders, cultural groups also include people differentiate by age, gender or religion. All psychological processes take place in a cultural context; traditionally that context in psychological research is North American and Western European college-educated white people. This limitation has seen its fair share of criticism in the overdependence on Western, Educated, Industrialized, Rich and Democratic (WEIRD) societies as samples in scientific research (Arnett, 2008; Henrich, Heine, & Norenzayan, 2010), CCSP looks to expand this. Beyond just including more groups, CCSP investigates if, how, and why different cultural contexts influence psychological processes.

It is a particularly important research topic today because social psychology research faces a replicability crisis. This crisis came out of work that showed that a significant majority of results in social psychology could not be replicated (Open Science Collaboration, 2015). From this crisis grew a debate that asked why this is the case and what we can do about it. Especially relevant is the spotlight on the importance of context or boundary conditions (Van Bavel, Mende-Siedlecki, Brady, & Reinero, 2016) - an important aspect of that being culture. From this, it is recognized that the field does not know the full extent of cultural context's influences on psychological processes or even how to measure cultural differences.

The challenges brought to light from the replicability crisis are helpful in understanding where cross-cultural psychology belongs in the framework of social psychology – namely, the focus on differences in replication environments and the question of applicability and replicability across the world. Therefore, cross-cultural psychology addresses the fundamental question: do our theories have boundary conditions that are culturally bound? Additionally, the cross-cultural perspective facilitates confronting the blind spots in the primarily western perspective of psychological concepts, which increase limitations in theory and approach (Chen, 2010).

The inherent value of cross-cultural psychology for social psychology as a whole is apparent but CCSP has some unique challenges compared to other sub-fields of social psychology that must be addressed. For example, because culture is embedded in specific human psychological contexts, there is difficulty establishing causal relationships. Furthermore, due to research conducted in different contexts there are more extensive problems with cross-cultural bias, which can broadly be defined as "any systematic source of distortion that challenges the validity of cross-cultural comparison" (Leung & van de Vijver, 2008, p. 148). Without addressing these issues, the field cannot adequately address relevant questions. As Leung and van de Vijver (2008) proclaimed "…causal inferences in cross cultural research are most convincing when supported by diverse evidence based on a sound theoretical basis, multiple source of data, different research methods, and explicit refutation of alternative interpretations" (p. 145).

In completing our understanding of CCSP we should differentiate it from other related approaches to the study of culture and psychology. *Cultural psychology* focuses on understanding the underlying meaning of cultural phenomena, while cross-cultural psychology focuses on causal inferences of culture (Leung & van de Vijver, 2008). *Indigenous psychology* identifies unique psychological processes without Western concepts as a starting point and assumes some lack of comparability, contrasting with cross-cultural psychology, which focuses on the comparison between groups. *Intercultural psychology* focuses on understanding the interaction between different groups in comparison contrary to CCSP that compares differences between groups (Smith, Fischer, Vignoles, & Bond, 2013).

In the next section, I will go over a brief history of the field of CCSP, followed by a review and critique of fundamental theories and conclude with a closer look at methodological challenges.

Brief history

Anthropological explorations of cultural comparisons can be traced back to early explorers (Hogg & Vaughan, 2011). Additionally, observed and recorded cultural differences through the approach of cultural psychology can be seen as early as the 20th century (Lonner, 2013; Smith et al., 2013). However, social psychologists avoided culture for some time because of their commitment to experimental methods and the view of cross-cultural research as being primarily descriptive (Hogg & Vaughan, 2011). This started to shift in the 1960's with the first publications of cross-cultural psychology journals and in 1972 with the inaugural conference for *International Association for Cross-Cultural Psychology*. According to Lonner (2013), a pioneer in the field, the development was as follows: "If the 1960s can be viewed as the seminal decade, the 1970s and 1980s can be considered a period that solidified cross-cultural psychology as a revolutionary development in the discipline." (p. 5).

For a window into the evolution of the field as a whole, Cretchley, Rooney and Gallois's (2010) systematic review of topics in the *Journal of Cross-Cultural Psychology* demonstrates the evolution of the journal over time. They reported constancy throughout the history of the publication with a strong emphasis on experimental psychology. As far as the progression of topics goes, in the early years there was a focus on child development, while in more recent years there has been an emphasis on aspects of culture such as values, orientation, and acculturation. This leads in turn to current times, where social psychology and personality are more central areas of publication.

In the early 1980's Hofstede brought wide attention to the field by systematically mapping countries in terms of psychological constructs. The relevance at the time was the emergence of international businesses and the lack of research on work culture; therefore, Hofstede filled the demand for guidance in this area (Søndergaard, 1994). To this day, organizational applications of cross-cultural differences remain a popular practical application. Because of the expansion of interest in cross-cultural studies and Hofstede's

work, especially in the area of social psychology, came efforts to investigate psychological mechanisms behind observed country-level differences. Markus & Kitayama (1991) exemplify this approach in their conceptualization of independent and interdependent self-construal as the mechanisms behind observed cultural differences. These concepts defined decades of cross-cultural research and therefore it is necessary to thoroughly understand these theories.

Problems with the concepts and theories in cross-cultural social psychology: focus on Individualism and Collectivism

One of the pillars of good research is having a strong theoretical argumentation. In CCSP, the theoretical framework on which the field is entrenched has primarily been analyzing the differences between individualism and collectivism, or independent versus interdependent self-construal, which were popularized by Hofstede (1980a, 1984), Triandis (1985, 1995) and Markus & Kitayama (1991) and others. Individualism and collectivism dichotomy also emerged as the primary framework because it fits the fact that most research in CCSP is a comparison of Western countries and East Asian countries (Cross, Hardin, & Gercek-Swing 2011; Oyserman, Coon, & Kemmelmeier, 2002). In this section we will explain more in-depth what the core theories entail as well as their limitations.

Hofstede's Cultural Dimensions

Hofstede's 1980 and 1984 books were highly influential in the field of CCSP and are often synonymous with CCSP itself (Kirkman, Lowe, & Gibson, 2006; Tsui, Nifadkar, & Ou, 2007). Currently, his 1984 book has over 54,900 citations, far and above anything else in cross-cultural research. Even when researchers are interested in other aspects of culture or not looking at his constructs directly, Hofstede's ranking is often used for selecting samples (Taras et al., 2014). Therefore, it is important to look at the role and contribution of his work, specifically pertaining to individualism and collectivism, which shapes the direction of cross-cultural research to this day.

In his own words, Hofstede describes his research as follows: “In the 1970s this author – more or less by accident – got access to a large survey database about values and related sentiments of people in over 50 countries around the world (Hofstede, 1980). These people worked in the local subsidiaries of one large multinational corporation: IBM. Most parts of the organization had been surveyed twice over a four-year interval, and the database contained more than 100,000 questionnaires.” (Hofstede, 2011, p. 6). From this data set, Hofstede factor analyzed a matrix of 32 value related questions from an initial 40 countries, and later 53 countries, and found values clustering around 4 dimensions: Individualism/Collectivism, Power Distance, Masculinity, and Uncertainty Avoidance. He later added two additional dimensions: Long/Short Term Orientation and Indulgence/Restraint (Hofstede, 2011). From this happenstance data, a profoundly influential theory of cultural difference was established.

Individualism and collectivism (IND-COL) garners more attention specifically because of its popularity and influence. As described by Taras et al., (2014), in a large meta-analysis of the effects of culture in the workplace, it was found that 88% of all effects in cross-cultural literature reported individualism and collectivism measures (Taras, Kirkman & Steel, 2010). Hofstede defined IND-COL as one dimension that measures “…the degree to which people in a society are integrated into groups” (Hofstede, 2011, p. 11). More specifically, individualism is “a loosely-knit social framework in which people are supposed to take care of themselves and of their immediate families only,” while collectivism “is characterized by a tight social framework in which people distinguish between ingroups and outgroups, they expect their ingroup to look after them, and in exchange for that they feel they owe absolute loyalty to it” (Hofstede, 1980b, p. 45). The measure was conceptualized as polar ends of a scale. Although Hofstede was not the first to use these categorizations, he popularized them in an explanatory role and is regarded as the father of IND-COL (Taras, Steel, & Kirkman, 2012).

Although extremely influential, there is also a lot of criticism of Hofstede's work (Kim & Sharkey, 1995; Levine et al., 2003; Voronov & Singer, 2002). First, in regards to the happenstance of the data: no cohesive theory existed before the collection of the data, only a series of post-hoc explanations which relied on factor analysis. This means that although the analysis could show which items belong together, it does not guarantee construct validity and is limited by the items in the original scale (Kagitcibasi, 1997). In his own defense, Hofstede claims this is not a problem due to extensive replications: "The number of external validations kept expanding, and the second edition of Culture's Consequences (Hofstede, 2001, Appendix 6, pp. 503-520) lists more than 400 significant correlations between the IBM-based scores and results of other studies." However, correlations to other studies do not necessarily mean that these other studies identified underlying factors as he described. Furthermore, other replication studies do not back up Hofstede's confidence in replications. For example, in Søndergaard's (1994) study analyzing 61 replications, only four replicated Hofstede's dimensions in their entirety and 15 showed partial confirmation. More recently in a meta-analysis, Taras, Kirkman and Steel (2010) found limited predictive values of Hofstede's four original dimensions, highlighting that personality traits and demographics were significantly better predictors of organizational outcomes. Hofstede defended any lack of replication by predicting that cultures will shift over time (Hofstede, 2011; Hofstede, Hofstede, & Minkov, 2010). However, this points to further criticism of the work, that it is not clear if and when, or how much culture changes (McSweeney, 2002). These theoretical weaknesses have contributed to the ongoing confusion for the role of Hofstede's work in cross-cultural psychology.

Although there are widespread criticisms as mentioned above, Hofstede's model continues to be the center of research that focuses on cultural issues. Many recently published articles highlight the way in which the vague theoretical understanding of the dimensions is an ongoing issue (e.g. Kotlaja, 2018; Reyes, 2018). Although Hofstede himself emphasizes

the limitations of the theory, especially in regard to the level of analysis, in that these are cultural level dimensions and not individual level ones, there remains much ambiguity in its theoretical application. In practice, IND-COL is often conflated with self-construal or individual level cultural differences, which defenders of Hofstede agree is not what he intends to be measuring (Hofstede, 2011; Hofstede, Hofstede, & Minkov, 2010). It is common for researchers to take parts of Hofstede's definition that fit their own hypothesis and ignore the parts that do not. For example, studies often use Hofstede's country level scores as an independent variable. However, the measures described by Hofstede do not match how the authors conceptualize the variable. Since cross-cultural research has become synonymous with IND-COL, it is used with increased frequency, which leads to even more usage overtime, while the research that follows is still not addressing the fact that the theoretical concepts are still lacking.

More recently, Minkov et. al. (2017) proposed a new "valid" measure of IND-COL; however, looking at the measures themselves, it is still not clear what IND-COL in fact mean. The Minkov et. al. (2017) publication describes IND-COL as a multifaceted dimension with conformism and social ascendancy aspects, with such items as "I am very religious – I rarely observe religious rules", "I respect all rules of my society – I decide which rules to respect" and "I would like to achieve fame – I see fame as useless". It is unclear how this is conceptually related to Hofstede's definition of IND-COL: "the degree to which people in a society are integrated into groups" (Hofstede 2011, p. 11) or the items from Hofstede's original scale, such as "have considerable freedom to adopt your own approach to the job" at the individualistic pole, and "have training opportunities (to improve your skills or learn new skills)" at the collectivist pole (Hofstede, 2010, p. 92-93). Therefore, we can conclude that Hofstede's IND-COL dimension is not very useful in that it is not clear what exactly is measured and what country rankings on this scale mean beyond a post-hoc explanation. The indiscriminate use of IND-COL means the terms lose their utility.

Self-Construal Theory

Following Hofstede was research on cross-cultural difference in the concept of self. Markus & Kitayama's (1991) independent and interdependent self-construal framework epitomizes this approach. An extremely popular foundational theory, it has over 21,300 citations to date. As defined by Markus and Kitayama (1991), independent self-construal is when "view of the self derives from a belief in the wholeness and uniqueness of each person's configuration of internal attributes" (p. 226) and interdependent self-construal is when "view of the self and the relationship between the self and others features the person not as separate from the social context but as more connected and less differentiated from others" (p. 227). Although Markus and Kitayama (1991) explicitly stated that these were only two of possibly many self-representations, self-construal became virtually synonymous with "how individuals see themselves in relation to others" (Cross et al., 2011 p. 143). Over the years there has been a plethora of research in this area on a wide variety of topics from cognition to affect to effects on motivation (see for a review: Cross, et al., 2011; Gudykunst & Lee, 2003; Smith, et al., 2013).

The relationship between Hofstede's IND-COL and independent and interdependent self-construal are most commonly described as follows: IND-COL represents the country level variable while self-construal represents the individual level variable (Cross et al., 2011). That is, people from a collectivist culture tend to be interdependent while those from an individualistic culture tend to be independent. However, there are still basic questions of how the individual level corresponds to the country level (Kagitcibasi, 1997). Other researchers in the field describe the relationship of independent and interdependent self-construal as facets of a larger cultural construct of IND-COL (Triandis, 1995).

There is a variety of measurement tools for self-construal, primarily those developed by Gudykunst and colleagues (1996), Kim and Leung (1997), Oyserman (1993), Singelis (1994), Takata (1993), and Triandis (1994). However, the extent of differences between these

scales is not clear and which, if any, best represent the concept at hand. Furthermore, results from research using these different scales tend to be inconsistent (Cross et al., 2011). For example, highlighting the problems, meta-analysis results showed that western people are not more individualistic, which is the core base of cross-cultural research theory. (Levine et al., 2003; Oyseman et al., 2002; Takano & Osaka 2018, Voronov & Singer 2002). This was startling to the field and some have declared that culture does not in fact reliably predict self-construal in the way the theory predicts (Matsumoto, 1999). For other researchers this was a catalyst for a new wave of questions. The literature has since reflected both on the problems and solutions to these measurement concerns.

The differentiation of the theoretical frameworks that emerged out of the criticism of Hofstede's original theory was that IND-COL is not a unidimensional construct but two orthogonal constructs (e.g. Markus & Kitayama, 1991; Oyserman, 1993; Singelis, 1994; Triandis, 1995). That is to say that independent self-construal and interdependent self-construal are unique constructs and therefore the degree to which an individual is higher or lower in one dimension is independent of the other. However, the literature has shown this discussion to be far from settled (Hardin, Leong, & Bhagwat, 2004; Taras et al., 2014; Vignoles et al., 2016). For example, Taras and colleagues (2014), in their observation of published works, explained that even though the concepts are considered orthogonal, the functional definitions and proceeding hypothesis are almost always paired opposites. Across six commonly used scales, 40 out of 45 attributes were postulated with an explicit opposite. Furthermore, in their meta-analysis of research that used these instruments, Taras and colleagues (2014) found that the dimensionality of the two concepts, looking at scales which describe both IND-COL and self-construal, depends on a number of factors: 1) the instrument used; 2) the characteristics of the sample, such as region; and 3) the level of analysis. This is very problematic because uncertainty about the relationship between IND and COL and

independent and interdependent self-construal leads to uncertainty in the interpretation of data as well as comparability of results.

Regarding limitations of the construct, researchers have tried to address solutions with various approaches. In the next sections I will outline some of those approaches. One approach is a theoretical clarification of the meaning of individualism and collectivism, another is a move away from the dual model of selfhood, and a third approach is that some of these discrepancies may be explained methodologically.

A promising area of inquiry to explain the discrepancy of findings is to understand the meaning of collectivism. Primary to the definition of collectivism is connection to one's group, but what is meant by the group? From this question a third dimension of self-construal was first proposed by Brewer and Gardner, (1996) and Kashima et al., (1995), that is the relational-self. This is the extent to which people define themselves in terms of close relational others, thus differentiating from the self, which is defined in terms of large groups and group membership. This differentiation was further highlighted by Brewer and Chen (2007). To approach the discrepancies in the literature and the mean of the collective, Brewer and Chen content-analyzed commonly used scales to categorize the content of the scale items. Two dimensions are specified – the first relates to elements of the items, that is, self-representations, beliefs, or values. The second is the target of the items, that is, who is the collective: the individual, a relational other, or a collective group? They found that clarifying these distinctions contributes to an understanding of the limitations of the interpretation of past research. Other researchers have also called to make this distinction and have found convincing results in areas such as gender difference in self-construal, and cross-cultural differences in intergroup trust (Cross, Bacon & Morris, 2000; Gabriel & Gardner, 1999; Yuki, Maddux, Brewer, & Takemura, 2005). Nonetheless, most research continues to focus on the more general dichotomous IND-COL and independent interdependent self-construal (Cross et al., 2011). Therefore, although this approach has potential, more research is needed.

Another way to tackle the limitations of the construct found in the literature outlined above is a more general look at the structure of the overall construct and if there are additional models of selfhood. Hofstede claimed that using six dimensions to describe cultures was the best way because it was not too few and not too many, and the dimensions used were all bi-polar (Hofstede, 2011). However, this does not necessarily reflect the reality of cultural differences. Triandis (1995) proposed that IND and COL are in fact multidimensional constructs, differentiating between vertical aspects of IND-COL that focus on hierarchies, and horizontal aspects of IND-COL that focus on egalitarian values. More recently, Vignoles and colleagues (2016) argue that the existing focus on independent versus interdependent self comes out of the focus on North American and East Asian samples; and that if we take a wider view of the world, we will find more diverse definitions of the self. With large samples from a diverse group of countries, they found a seven-dimensional model of selfhood. However, this work was limited in that questions were derived from existing measures. Furthermore, this work needs to be expanded to understand the implication of multiple dimensions of selfhood rather than IND-COL dichotomy.

Regarding where the theoretical construct of IND-COL and self-construal fit overall in cross-cultural research, I think there is utility in the concepts and there is a lot of past research with compelling findings that have enriched our understanding of cross-cultural differences. However, there is also a dearth of replication and a glut of research that perpetuates a lack of clarification. The theoretical formulation as of now is a large umbrella that everything fits under. However, the conceptualization in research involves picking and choosing relevant concepts, which will not lead to strong theory building. Instead, we need to work on refining our definition and replicating past findings, thereby clarifying concepts resulting in more concise theories. In addition, we need to be sure that our methods are sound. This is an equally important aspect which I will discuss in the next section.

Problems with Measurement

Another pillar to sound research practice is firm research methods; as is clear from the above review, this has also been a problem with IND-COL research. However, the research method limitations go beyond the definition and measurement of one scale. In general, CCSP reflects social psychology's overall methodological concerns; therefore, methodological challenges of cross-cultural research are similar to those in all research, with regards to internal and external validity, etc. However, there are some aspects that are particularly challenging for CCSP, as successful CCSP goes beyond merely using the same materials to collect data in another context.

Particularly, issues in cross-cultural research are those related to biases in both sampling and measurement equivalence. Bias in general is when an outcome variable does not correspond to the concept one wants to measure. Although all areas of psychology need to be aware of the ramifications of sampling, this can be particularly precarious when comparing cross-national groups because of unintended confounding factors. For example, using convenience samples of college students in developed countries and developing countries might not be as comparable as a research team expects because university attendance might be correlated to wealth in different ways in different countries (Cohen, 2007). Another issue is that otherwise comparable cultural groups might have different familiarity with research methods, therefore resulting in differences due to measurement errors and not the questions at hand.

The idea of comparability is the foundation of CCSP; however, we must question if we are really comparing the same thing across different groups since the lack of measurement equivalence can produce methodological artifacts. Therefore, it is important to take steps to insure comparability both in the preparation of a study and in verifying the statistical comparability after the fact.

A large source of potential methodological issues is with the translation process. Language is a very prominent feature of culture and translation often results in a transformation of meaning. The standard translation-back translation (Brislin, 1970) process is considered the best practice to minimize translation error; however, this is not foolproof. Direct translations also need to be conceptually comparable. An increasingly discussed hot topic among cross-cultural researches is if translations-back translations by a native speak who is naive to the construct really captures a construct or if there is a better way.

In addition, regarding assessment of comparability in CCSP, response styles is an important aspect to consider. Response styles are systematic tendencies in the manner in which one responds to rating scales, in that observed scores will be unrelated to the true scores of the individual (Fischer, 2004). Some types of response styles are modest or extreme responding, such as acquiescence, which is selecting extreme answers, shifting responses to either end of the scale (Fischer, 2004). In cross-cultural research, these response styles can be seen as error artifacts or as aspects of the cultural differences at hand (Smith, 2011). For researchers who view response styles as method bias in cross-cultural studies, it is common to follow standardized ipsatization procedures as a way of controlling for country level acquiescence. This has been the approach in many large scale cross-cultural research projects (i.e. Gelfand, 2011; Hofstede, 2001; House et al., 2004; Schwartz, 2004). However, this approach is insufficient because not only does it require excluding potentially relevant information, but it is also statistically problematic (Cheung & Rensvold, 2000; Fischer, 2004).

Therefore, it is important to test for measurement equivalence rather than depend on standardization procedures. Testing for measurement invariance, which more precisely, evaluates the comparability of scales across heterogonous groups is increasingly acknowledged as a necessary part of cross-cultural research (Cheung and Rensvold, 2000). However, this is still not universally done, in that studies show only 17% of cross-cultural research studies test measurement equivalence (Boer, Hanke, & He, 2018; Chen, 2008).

Invariance testing is still not widespread because of inadequate education in execution of invariance testing, and because finding invariance by traditional means is often unattainable for many groups (Byrne & van de Vijver, 2017; Steenkamp & Baumgartner, 1998).

Beyond these general challenges for equivalence in the literature, there are also specific methodological responses to the limitations and unexpected findings of IND-COL literature. For example, Heine, Lehman, Peng, and Greenholtz (2002) focus on the issues with the Likert scale, in that the lack of expected cross-cultural differences is due to references group effects. Survey questions require participants to reflect on the question with a degree of judgment. For example, a question asking the participant if a statement describes them "very much" or "very little" on a continuum requires a reference point. The most likely reference point will be to compare to similar others (Heine et al., 2002). For example, if participants in Japan are asked if they are tall, they will compare themselves to other Japanese people and give a different answer than if explicitly asked to compare themselves to American people. With this reasoning, Heine and colleagues (2002) found that when you manipulate the reference group to explicitly reference other cultures, there is an enhanced effect of expected cultural differences with regards to IND-COL. Also, He, Buchholz, and Klieme (2017) found that by using anchoring vignettes as a means to reflect reference group effects, there was increased comparability and predictive validity for a large scale multi-nation study.

Current work

From this general review of the literature, it is clear that there are many areas of research that we need to further explore. The following chapters' record experimental research that tackles specific projects related to the broad challenges outlined above.

Part one will focus on self-construal and how we define our groups. As discussed above in cross-cultural social psychology, the understanding of the self with Hofstede's simplistic bi-polar dimension is limited in its utility, and other approaches, while promising, need more refinement and clarification. Although specification of the relational self has been

acknowledged as distinct since the 1990s, the tools of measurement and theory continue to underutilize the distinction (Cross et al., 2011). Therefore, in this section will take on the approach of identifying the differences between the collective self: the relational self as defined by relationships to close others and the collective or categorical self defined by connection to categorical groups. Across two studies, I will look at similarities and differences in how the in-group is defined both within and between cultures and how that is reflected in observed victimization. Chapter 2 reports on a vignette study which seeks to experimentally differentiate the relational self and group collective self by looking at the emotional reaction to another person's injustice in three cultures. This study clarifies ambiguity in the research in regard to cultural similarities or differences in relative importance of self-representations. This design improves on past studies by clearly distinguishing between the relational self and the group collective self for which there is often overlap. Further, this study uses a culturally neutral task in order to account for other additional cultural influences apparent in past research.

Chapter 3 also investigates the question of differences between groups and how they are incorporated in the self in the context of others injustice. In contrast to past research, this study uses novel methods to focuses on within-culture differences in measured and manipulated self-construal. Regardless of overall cultural norms, there are individual and situational differences in the relative importance of self-construal, this study aims to look specifically at those difference within one non-Western country. Further, this study can contribute of the clarity of connection between relational self and categorical/collective self and types of groups' membership, again distinguishing between the possible overlap between group types, using minimal group paradigms.

Part two will take on methodical challenges, in both understanding the depths of those challenges and proposing potential solutions: Firstly, by investigating a different aspect of response style that has previously been neglected in cross-cultural research, namely cultural

difference in careless responding; and secondly, by looking at limitations and new methods of measurement invariance.

Although researchers have taken a variety of approaches to methodological differences in conducting cross-culture studies, very few have looked at data quality, mode of data collection and culture together. Although there is some acknowledgement of the influences of variance in familiarity of taking part in studies (Cohen, 2007), this has not been systematically investigated. Chapter 4 fills this gap by brings attention to the problem in differences in data quality. Differences in data quality may have influenced the proceeding literature in unseen ways, especially in the sense that much of our understanding of cross-cultural difference comes from self repost surveys that give an inaccurate comparison if there is cross-cultural variability in responding carelessly. Further, with the emergences of increasingly diversity in the use of technology in data collection, the role of differences between modes should be reexamined especially across different cultures. This study further contributes to the literature by taking into account not only cross-national cultural groups but also different age groups by looking at students in Germany and both students and adults in China.

Finally, Chapter 5 further builds on the fundamental methodological issues concerned with the literature's focus on survey data. This study does a secondary data analysis of data collected across eight countries and eight measures to compare traditional and new approaches to establishing measurement comparability. Necessary for any cross-cultural research is the comparability of the materials; however in practice this is not always done. Therefore this investigation is vital in expanding our understanding on new methods and how these new methods relate to previous ones.

References

Arnett, J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63, 602-614.

Boer, D., Hanke, K., & He, J. (2018). On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. *Journal of Cross-Cultural Psychology*, 49, 713-734.

Brewer, M. B., & Chen, Y. R. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological review, 114*, 133-151.

Brewer, M. B., & Gardner, W. (1996). Who is this "We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71*, 83–93.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, *1*, 185-216.

Byrne, B. M., & van de Vijve, F. J. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema*, *29*, 539-551.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of personality and social psychology, 95*, 1005-1018.

Chen, S. X. (2010). From emic to etic: Exporting indigenous constructs. *Social and Personality Psychology Compass, 4*, 364-378.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescent response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*, 187-212.

Cohen, D. (2007). Methods in cultural psychology. *Handbook of cultural psychology*, 196-236.

Cretchley, J., Rooney, D., & Gallois, C. (2010). Mapping a 40-year history with Leximancer: Themes and concepts in the Journal of Cross-Cultural Psychology. *Journal of Cross-Cultural Psychology, 41,* 318-328.

Cross, S. E., Bacon, P. L., & Morris, M. L. (2000). The relational-interdependent self-construal and relationships. *Journal of personality and social psychology*, 78, 791-808.

Cross, S. E., Hardin, E. E., & Gercek-Swing, B. (2011). The what, how, why, and where of self-construal. *Personality and Social Psychology Review, 15,* 142-179.

Fischer, R. (2004). Standardization to account for cross-cultural response bias: A classification of score adjustment procedures and review of research in JCCP. *Journal of Cross-Cultural Psychology, 35,* 263-282.

Gabriel, S., & Gardner, W. L. (1999). Are there "his" and "hers" types of interdependence? The implications of gender differences in collective versus relational interdependence for affect, behavior, and cognition. *Journal of Personality and Social Psychology, 77,* 642-655.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., ... & Aycan, Z. (2011). Differences between tight and loose cultures: A 33-nation study. *Science, 332,* 1100-1104.

Gudykunst, W. B., & Lee, C. M. (2003). Assessing the validity of self-construal scales: A response to Levine et al. *Human Communication Research, 29,* 253–274.

Gudykunst, W. G., Matsumoto, Y., Ting-Toomey, S., & Nishida, T. (1996). The influence of cultural individualism-collectivism, self-construals, and individual values on communication styles across cultures. *Human Communication Research, 22,* 510-543.

Hardin, E. E., Leong, F. T., & Bhagwat, A. A. (2004). Factor structure of the self-construal scale revisited: Implications for the multidimensionality of self-construal. *Journal of Cross-Cultural Psychology*, 35, 327-345.

He, J., Buchholz, J., & Klieme, E. (2017). Effects of anchoring vignettes on comparability and predictive validity of student self-reports in 64 cultures. *Journal of Cross-Cultural Psychology, 48*, 319-334.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of personality and social psychology, 82*, 903-918.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences, 33*, 61-83.

Hofstede, G. (1980a). *Culture's Consequences: International Differences in Work-Related Values.* Beverly Hills, CA: Sage.

Hofstede, G. (1980b). Motivation, leadership, and organization: Do American theories apply abroad? *Organizational Dynamics, 9*, 42-63.

Hofstede, G. (1984) *Culture's consequences: International differences in work-related values.* London: Sage.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online readings in psychology and culture, Article 8.*

Hofstede, G., Hofstede, G. J. & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind* (Rev. 3rd ed.). New York: McGraw-Hill.

Hogg, M., & Vaughan, G. (2011). Social Psychology. Sixth edition. Pearsons.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage publications.

Kagitcibasi, C. (1997). Individualism and collectivism. In J. W. Berry, M. H. Segall, & C. Kagitcibasi (Eds.), *Handbook of cross-cultural psychology: Vol. 3. Social behavior and applications* (pp. 1–49). Boston: Allyn and Bacon.

Kim, M. S., & Sharkey, W. F. (1995). Independent and interdependent construals of self: Explaining cultural patterns of interpersonal communication in multicultural organizational settings. *Communication Quarterly, 43,* 20-38.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of culture's consequences: A review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies*, 37, 285-320.

Kotlaja, M. M. (2018). Cultural contexts of individualism vs. collectivism: Exploring the relationships between family bonding, supervision and deviance. *European Journal of Criminology*, 1-18.

Lehman, D. R., Chiu, C. Y., & Schaller, M. (2004). Psychology and culture. *Annu. Rev. Psychol.*, *55*, 689-714.

Lehman, D. R., Chiu, C. Y., & Schaller, M. (2004). Psychology and culture. *Annu. Rev. Psychol.*, *55*, 689-714.

Leung, K., & Van De Vijver, F. J. (2008). Strategies for strengthening causal inferences in cross cultural research: The consilience approach. *International Journal of Cross Cultural Management, 8,* 145-169.

Levine, T. R., Bresnahan, M. J., Park, H. S., Lapinski, M. K., Wittenbaum, G. M., Shearman, S. M., . . Ohashi, R. (2003). Self-construal scales lack validity. Human *Communication Research, 29,* 210-252.

Lonner, W. (2013). Chronological Benchmarks in Cross-Cultural Psychology. Foreword to the Encyclopedia of Cross-Cultural Psychology. *Online Readings in Psychology and Culture*, 1.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98,* 224-253.

Matsumoto, D. (1999). Culture and self: An empirical assessment of Markus and Kitayama's theory of independent and interdependent self-construal. *Asian Journal of Social Psychology, 2*, 289-310.

Matsumoto, D. & Juang, L. 2008. *Culture and psychology*, 4, Belmont, CA: Wadsworth.

McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith-a failure of analysis. *Human relations, 55*, 89-118.

Minkov, M., Dutt, P., Schachner, M., Morales, O., Sanchez, C., Jandosova, J., ... & Mudd, B. (2017). A revision of Hofstede's individualism-collectivism dimension: A new national index from a 56-country study. *Cross Cultural & Strategic Management, 24*, 386-404.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251).

Oyserman, D. (1993). The lens of personhood: Viewing the self and others in a multicultural society. *Journal of Personality and Social Psychology, 65*, 993-1009.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 3–72.

Reyes, M. L. (2018). Social mobility attributions in East Asian and Pacific cultures: Power distance and individualism as moderators of self-attribution bias. *Journal of Pacific Rim Psychology, 12*.

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin, 20,* 580-591.

Smith, P. B. (2011). Communication styles as dimensions of national culture. *Journal of Cross-Cultural Psychology, 42,* 216-233.

Smith, P. B., Fischer, R., Vignoles, V. L., & Bond, M. H. (2013). *Understanding social psychology across cultures: Engaging with others in a changing world*. London, UK: Sage.

Søndergaard, M. (1994). Research note: Hofstede's consequences: a study of reviews, citations and replications. *Organization studies, 15*, 447-456.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78-90.

Schwartz, S. H. (2004), *Mapping and interpreting cultural differences around the world.* In Vinken, H., Soeters, J. and Ester P. (eds.), Comparing cultures, Dimensions of culture in a comparative perspective. Leiden, The Netherlands: Brill.

Takano, Y., & Osaka, E. (2018). Comparing Japan and the United States on individualism/collectivism: A follow-up review. *Asian Journal of Social Psychology, 21*, 301-316.

Takata, T. (1993). Social comparison and formation of self-concept in adolescent: Some findings about Japanese college students. *Japanese Journal of Educational Psychology, 41*, 339-348.

Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of culture's consequences: A threedecade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology, 95*, 405-439.

Taras, V., Sarala, R., Muchinsky, P., Kemmelmeier, M., Singelis, T. M., Avsec, A., ... & Hardin, E. E. (2014). Opposite ends of the same stick? Multi-method test of the dimensionality of individualism and collectivism. *Journal of Cross-Cultural Psychology, 45*, 213-245.

Taras, V., Steel, P., & Kirkman, B. L. (2012). The times they are A-Changin': Improving cultural indices and rankings based on a meta-analysis of Hofstede's dimensions. *Journal of World Business, 47,* 329-341.

Triandis, H. C. (1994). INDCOL (Unpublished research scale on individualism and collectivism). Champaign: University of Illinois.

Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.

Triandis, H. C., Leung, K., Villareal, M. J., & Clack, F. I. (1985). Allocentric versus idiocentric tendencies: Convergent and discriminant validation. *Journal of Research in personality, 19,* 395-415.

Tsui, A. S., Nifadkar, S. S., & Ou, A. Y. (2007). Cross-national, cross-cultural organizational behavior research: Advances, gaps, and recommendations. *Journal of Management, 33,* 426-478.

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences, 113,* 6454-6459.

van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology, 67,* 2-12.

Vignoles, V. L., Owe, E., Becker, M., Smith, P. B., Easterbrook, M. J., Brown, R., ... & Lay, S. (2016). Beyond the 'east–west' dichotomy: Global variation in cultural models of selfhood. *Journal of Experimental Psychology: General, 145*, 966-1000.

Voronov, M., & Singer, J. A. (2002). The myth of individualism-collectivism: A critical review. *Journal of Social Psychology, 142,* 461-480.

Yuki, M., Maddux, W. W., Brewer, M. B., & Takemura, K. (2005). Cross-cultural differences in relationship- and group-based trust. *Personality and Social Psychology Bulletin*, *31*, 48-62.

Kashima, Y., Yamaguchi, S., Kim, U., Choi, S., Gelfand, M. J., & Yuki, M. (1995). Culture, gender, and self: A perspective from individualism-collectivism research. *Journal of Personality and Social Psychology, 69,* 925-937

Part 1: Theory

II. Chapter 2: Relational and Group Collective Self Responses to Observed Victimization Across Cultures

Zoe Magraw-Mickelson and Mario Gollwitzer

Abstract

Mental representations of the "self" consist of both individual aspects (i.e., how one differs from other people) and collective aspects (i.e., how one relates to other people), with collective aspects further consisting of interpersonal relations (the "relational" self) and of memberships in social groups (the "group collective" self). Some researchers assume that there is a universal motivational hierarchy in self-representations (with the relational self being more relevant than the group collective self). Other research suggests that the relative importance of self-representations varies across cultures. This paper tests the motivational hierarchy hypothesis in a cross-cultural context. Emotional reactions (anger, outrage, vengeful intentions) to observed victimization of a collective or relational group member were assessed in Germany, Japan, and the USA. In line with the motivational hierarchy hypothesis, we found, across all three countries, evidence for the primacy of the relational self over the group collective self.

Introduction

Say you observe an instance of victimization, for example, an individual is treated unjustly by a third party. How would you feel? In general, most people would be uncomfortable with such a situation and believe it is unjust. However, would our degree of anger or moral outrage depend on our connection to the victimized individual? Would it make a difference whether this individual belongs to our in-group or an out-group? Would it matter if we knew the victim personally? To answer these questions, we must look to how the self is defined and the literature that differentiates between ways in which others are included in the self-concept. Predictions from the "hierarchy of the self" literature would say we always have a stronger emotional reaction when the victim is someone within our social network. However, cross-cultural research would predict there are cultural differences depending on the primacy of differing self-concepts.

Cultural Differences in Self-Representations

A recurring theme in cross-cultural psychology is the core definition of the self and where its boundaries lie. Broadly speaking, much theory is built on the assumption that Western cultures have a self-concept based on individual autonomy, while Eastern cultures have a self-concept based on social connection and interdependence with in-group members (Markus & Kitayama, 1991). These concepts have fallen under various labels (e.g., individualism vs. collectivism, independent vs interdependent self-construal) that have been both helpful and limiting (Oyserman, Coon, & Kemmelmeier, 2002). Two key areas necessary for further investigation are: 1) what is meant by "the collective" or "the group"? and 2) how do the different selves manifest within an individual? Understanding the nuances of group self-concepts, which is often examined in the context of cross-cultural comparisons, can enhance and complement looking at cross-cultural similarities. This, in turn, can clarify expected similarity or differences in relation to observed injustice.

Building on Markus and Kitayama's (1991) independent vs interdependent self-

construal dichotomy, Brewer and Gardner (1996) presented the concept of, and preliminary evidence for, three different self-representations, formulating that the self includes (1) our autonomous self as an individual ("individual self-representation"), (2) our personal relationships with other individuals ("relational self-representation"), and (3) our formal memberships in social groups ("group collective self-representation"). Each self represents different levels of inclusiveness. The latter two selves can be differentiated in that the *relational* self consists of social connections: this self-representation includes people one personally knows (e.g., friends, colleagues, fellow students). The *group collective* self consists of social categories one belongs to: this self-representation includes people who also belong to the same categories (e.g., people working in the same company, studying at the same university, living in the same country), but who one does not necessarily know personally.

This general concept leads to two lines of distinct but related questions. First, do these concepts help us understand cross-cultural differences in social behaviors? Second, is there a "hierarchy of selves," that is, is one self more important than others? Past work has taken these two questions into consideration but has not sufficiently integrated these issues. In the present study, we test the motivational hierarchy hypothesis in a cross-cultural context. This is important because cross-cultural research adds contextual information to the examination of basic processes that are often overlooked. Our study is not the first to test the motivational hierarchy hypothesis in different cultural contexts, but, as we will explain below, our methodological approach is more convincing than previous studies in this area, in that potential overlap of the concepts is minimized. Furthermore, our study is applied in a setting that tests the expanding of self-oriented emotions to include others in the context of injustice.

Relational and Group Collectivism

Traditionally, cross-cultural research has contrasted individualistic with collectivistic societies – which have been defined as societies in which members' self-concept is

interconnected with the group (Triandis, 1995). This definition has been criticized for being too narrow and simplistic. More specifically, Oyserman et al., (2002) have argued that the term "collectivistic" might be too general and not always useful in our understanding of categorizing individuals and cultures. Following Brewer and Gardner's (1996) descriptions of the three levels of self-representation (individual, relational, and group collective; see above), Brewer and Chen (2007) applied a concrete differentiation between the latter two. More precisely, they suggested dividing collectivism into "relational" and "group" collectivism – *relational collectivism* being the connection to a group of people with whom one has a relationship on a personal level (such as a family member, friend, or colleague) and *group collectivism* being the connection to the larger collective group with whom one shares a social categorization (such as someone from the same city or country; cf. Brewer & Gardner, 1996; Yuki, 2003). It is important to note that the group collective self includes a perceived connection with an individual person as well as with the group as a whole (Brewer & Gardner, 1996).

Therefore, perceived cultural differences between collectivistic and individualistic cultures may be due to the dissimilarity in the terms by which others are included in the self and how this dissimilarity contributes to the definition of the self. More specifically, a relational self-representation is based on interpersonal connections (e.g., "I am a colleague of Dr. Brewer and Dr. Gardner") whereas a group collective self-representation is based on memberships in social categories (e.g., "I am a member of the psychology department"). Notably, these self-representations also predict how individuals behave towards others, for instance, how they perceive and evaluate other people's actions. Consequently, group memberships (i.e., whether the judged individual belongs to one's in-group or to an out-group) are more relevant in "group collectivistic" cultures, whereas personal connections (i.e., whether they know somebody personally or not) are more relevant in "relational collectivistic" cultures.

Evidence for this conceptualization of relational and group collectivism comes from looking at trust and group category membership (Realo & Allik, 2009; Yuki, Maddux, Brewer, & Takemura, 2005), allocation task in nested groups (Lee, Adair, Mannix, & Kim, 2012), and the norm of reciprocity (Chen, Chen, & Portnoy, 2009). For instance, Yuki et al. (2005) found that US-American students were more inclined to trust other students from their own university (i.e., their in-group) than students from another university (i.e., an out-group), whereas Japanese students were more inclined to trust students from another university more if they knew somebody from that university personally (compared to students from a university where they did not know someone personally). In other words, group membership was more important to American respondents, who tend to have a strong categorically-defined ("group collectivistic") self-representation, whereas the possibility of a personal relationship was more important to Japanese respondents, who tend to have a strong relational self-representation (Brewer & Chen, 2007; Yuki et al., 2005).

Hierarchy of the Selves

These different selves exist across all cultures, but cultures may vary in the relative endorsement of the three. As originally conceptualized by Markus and Kitayama (1991), people from East Asian cultures would embrace the relational self and people from Western cultures would be more strongly inclined toward the individual self, while Brewer and Chen (2007) argue that people from Western cultures would also have a stronger group collective self. However, Gaertner and colleagues (Gaertner, Sedikides, & Graetz, 1999; Gaertner et al., 2012; Gaertner, Sedikides, Vevea, & Iuzzini, 2002) took a different conceptual approach. They acknowledge all three self-representations as important facets of identity, but argue that these self-definitions are hierarchically structured – regardless of one's cultural background. Across many studies they found evidence for the primacy of the individual over the relational self, and for the relational over the group collective self (Gaertner et al., 1999, 2002, 2012). In these studies, participants were presented with a threat to the self as an individual, to the

relational, or to the group collective self. For example, in earlier studies that compared the individual self to the group collective self (Gaertner et al., 1999), university students were told that they (individual self) or their university (group collective self) had performed poorly in a creativity test, and measured the reaction to this threat through a variety of measures such as anger, mood, and belief in the truthfulness of the feedback. They found that participants had a stronger adverse reaction to the threat to the individual self compared to the group collective self. With operationalizing the three different selves that way, Gaertner and colleagues consistently found that people react more strongly to situations that imply a threat to their individual self, followed by a threat to their relational self, and, lastly, a threat to their group collective self. For example, in a threat avoidance paradigm, participants avoided writing about ill fate more when it befell the self than when it befell a person they shared a personal relationship with (relational self), followed by a condition in which it befell a member of their most important group (group collective self) (Gaertner et al., 2012, Study 2).

In real life, there is often an overlap between the relational and the collective self. For instance, a person I know (i.e., relational self) can also belong to the same company (i.e., collective self). It makes sense to assume that reactions to situations that target the relational and the collective self are positively correlated when looking at individuals who are relevant for a person's self-representation. Research trying to prove that the relational self is more universally relevant than the group collective self needs to work hard to clearly disentangle the two selves. Here, we argue that there is room for improvement in the designs used by Gaertner and colleagues in this regard. For instance, learning that a friend of mine failed in a creativity test might make me more aroused than learning that a fellow student of mine failed, but it is hard to say how exactly this difference in arousal can be explained as long as it is unclear whether the "friend" belongs to an in- or an out-group, and whether I actually have a personal connection to the "fellow student" or not. This is what we aim to achieve with the present research.

Further, Gaertner et al. (2012) looked at cultural differences in two of their studies. These studies included samples from both Western countries (UK or USA) and East Asia (China). Although these studies found some evidence for a pan-cultural motivational hierarchy of the selves, the results were not definitive in the ability to discount the role of the cultural context in both methodology and in the results. For example, the operationalization used with the cross-cultural samples measured areas in which an individual has more personal control over the outcome and in this way has more direct agency, which favors the individual self and disfavors the group collective self. That is, when asked about future goals in reality we have more control over our own personal action and are therefore more likely to think of goals for which we can directly influence the outcome.

Beyond these limitations, the authors themselves called for future research to test their model in different contexts and with different populations. This is precisely what the present paper intends to accomplish. We will apply the questions of self representation primacy to the opening situation: how does the primacy of the relational and group collective self influence one's reaction to perceived injustice? Gaertner and colleagues acknowledge past research in the area of cross-cultural psychology; however, the focus on similarity between cultures leaves out the specificity of the concepts, which are important in cross-cultural contexts. That is, the hierarchy established in the research outlined above may depend on the overlap of the relational and group collective self-representation.

The Present Research

The current study aims to reexamine the hierarchy of the self under more stringent definitions of the relational and group collective self. We crossed relational connection, defined as someone with whom the participant has a social tie or who is part of their interpersonal network, with group collective connection, defined as someone with whom one shares category membership. Manipulating relational and group collective connection orthogonally is particularly beneficial because we can distinctly measure the unique effects of

different types of relationships. Gaertner et al. (2012) also acknowledge the role of overlapping self-representations, and investigated the overlap between the individual self and the relational self, and between the individual self and the collective self. However, there is reason for further investigation because they did not account for the overlap between the relational and collective self. Without distinguishing these two overlapping selves, it is difficult to say that one is in fact primary to the other.

In addition, the measurement of vicarious anger with specific people rather than a group as a whole was necessary to keep "construal level" constant. Drawing from construal level theory, objects or events can be represented more abstractly or more concretely, with differences in the level of abstraction being associated with differing judgments and behaviors (see McCrea, Wieber, & Myers, 2012). A group as a whole would be represented on a higher construal level than a specific individual. Therefore, comparing the observed victimization of a group would evoke different emotional reactions compared to the victimization of a singular relational other. The choice to represent the categorical group with an individual exemplar allowed us to compare the self-representations without confounding the level of construal.

Furthermore, this research addresses the question of the role of culture, which is still not conclusively answered. For instance, in Study 4 of Gaertner et al. (2012), the results fit the hierarchy of the self with a notable exception: American males showed no difference between relational and group collective self. This could fit a cultural context explanation as past studies showed that people from East Asian cultures and women are more interdependent (Cross & Madson, 1997). Therefore, the results are in line with the idea that women and Chinese male participants would show primacy of the relational self over the group collective self, and that this difference would diminish for Western men.

We tested the motivational hierarchy hypothesis by looking at emotional reactions to an observed victimization of others. Our central dependent variable was the extent to which participants experienced vicarious anger and moral outrage toward what happened to the other

person. We decided to investigate participants' emotional reactions to observed injustice because research suggests that proximity – both physical and symbolic – as well as in-group membership and similarity are important for sympathy and the desire to help victims (Loewenstein & Small, 2007). This is supported by neuropsychological evidence that shows through fMRI that the activation of regions associated with one's own pain mirrors the emphatic response toward the pain of a friend, but not a stranger (Beckes Coan, & Hasselmo, 2013). Further, work looking at justice sensitivity has shown that the sensitivity to feeling like a victim is extended beyond the self to include in-group members (Süssenbach & Gollwitzer, 2015). Injustice befalling others can also threaten the self if those others are relevant to the individual.

Participants were sampled in three different cultures: Germany, the USA, and Japan. In the literature, Germany and the USA are categorized as more individualistic culturally, while Japan is categorized as more collectivistic (Hofstede, 1980, 2001). This allows us to reexamine the contextual primacy hypothesis which is predicted by Brewer and Chen (2007), and was disregarded by Gaertner et al. (2012).

To control for individual differences that may influence the response to instances of injustice, we included a series of other measures. First, justice sensitivity from the victim's perspective ("victim sensitivity") has been shown to affect one's anger/outrage toward perceived injustice at one's own disadvantage (Gollwitzer, Schmitt, Schalke, Maes, & Baer, 2005; Schmitt, Gollwitzer, Maes, & Arbach, 2005; Süssenbach & Gollwitzer, 2015). Next, not all people within a culture adhere to cultural norms to the same degree. Therefore, it is also important to measure individual differences in the endorsement of the three self-representations (i.e., individual, relational, group; Leung & Cohen, 2011). Finally, because the situations in the vignettes are intended to invoke participants' reactions to harm done to one's in-group, the degree to which participants identify with the in-group in question is measured.

Methods

Participants

Participants were sampled in three different countries. In Germany (N = 273, 64% female), a university employee list-serve was used to recruit participants, and compensation was entry in a lottery for a tablet computer. The American (N = 214, 45% female) sample was recruited through the Crowdflower web platform for which participants received $1.00 for participating. The Japanese (N = 309, 54% female) sample was recruited through a survey service provider, and compensation paid through the service. The ages of German participants ranged between 19 and 65 years (M_{Age} = 36.08, SD_{Age} = 11.42); for American participants, the ages ranged between 19 and 73 years (M_{Age} = 33.49, SD_{Age} = 10.62); and for Japanese participants, the ages ranged between 20 and 75 years (M_{Age} = 45.54, SD_{Age} = 14.13).[1]

Procedure

After a welcome message, participants were asked to carefully read a series of four vignettes which depicted a workplace conflict situation. Each vignette was followed by a page with two manipulation check items, and another page with items measuring the dependent variable in this study, vicarious anger and outrage. This was followed by the next vignette situation. The names of the perpetrator, the target and the specific victimization differed across the four vignettes. The order in which the four vignettes were presented was held constant for all participants; however, the order in which the four experimental conditions appeared was counterbalanced (Greek-Latin square design). Furthermore, participants were randomly assigned to a male or female version in which both the perpetrator's and target's gender were matched (either both male or both female). Following the vignettes and dependent variables, participants completed the victim sensitivity, the self-representations,

[1] The Japanese and American versions of the survey included two bogus questions, and participants who missed either question were excluded from analysis (American sample: N = 123; the Japanese sample number of excluded participants was unreported by the data collection company). The exclusion criteria for the German sample were based on the manipulation check items described below.

and the in-group identification measures. Finally, a number of demographic variables (i.e., age, gender, employment status, size of employer, work experience, and type of industry[2]) were assessed. After that, participants were thanked and directed to their reward.

Vignettes

All materials were first written in English and included rounds of translation and back translation to each of the respective languages[3]. Each vignette confronted participants with a situation describing specific instances of victimization. Each participant read four victimization vignettes (due to the relational connection × group collective connection within-subject design). For example, victimization in the "no relational connection / group collective connection" condition was:

"Imagine you work for a large company. Your boss (let's call him Mr. Parris) calls in someone from your department *that* you don't know personally *(let's call him John) to his office to discuss their time sheet. There were days in the past month that John worked from home. However, Mr. Parris argues that John didn't have permission to do so, so he discounts some of John's work hours. Mr. Parris was aware that many other colleagues also work from home in the past month, but they were not reprimanded."* (see Appendix A for the full text of all the vignettes).

Manipulation Checks

Following each vignette there were two manipulation check items. In the German sample, this took the form of yes/no questions (e.g., "Was John (a) from the same department as you or (b) from a different department than you?" and "Do you know John personally? (a)

[2] The demographic variables did not affect the findings and are not reported on further.

[3] First, all materials were translated by researcher assistants naïve to the materials from English to the target language (German or Japanese) and then back to English by another naïve translator, following the established back translational procedure (Brislin, 1970). The back-translated English and the original English materials were then compared for discrepancies. These instances were then discussed and the entire document was checked for readability by two native speakers familiar with the materials. Next, a second back translation by a naïve translator and a check for any further discrepancies followed. If major changes were still needed, the materials were again discussed by the committee. If the two English translations were deemed equivalent, the translation was ready for use.

yes (b) no"). Participants answered 8 of these questions (2 items for each of the 4 vignettes). 193 participants (68.2 %) answered all 8 manipulation check questions correctly, 81 participants (28.6%) missed one or two questions, and 10 participants (3.2%) answered 3 or more questions incorrectly and were thus excluded from further analyses.

American and Japanese participants answered the manipulation check questions on a 6-point Likert type scale (e.g., "To what extent do you feel that John and you belong to the same group, the same 'category'?" from 1 = "not at all" to 6 = "very much;" and "To what extent do you feel personally connected to John?" from 1 = "totally unconnected" to 6 = "strongly connected"). For the American sample, in the vignette conditions in which participants were told they had a relational connection with the victim (M = 4.14, SD = 1.12), participants felt significantly more connected than when there was no relational connection (M = 3.58, SD = 1.28), $t(213) = -5.99$, $p < .001$, $d = 0.41$. In the vignette conditions in which participants were told they belonged to the same group as the victim (M = 4.15, SD = 1.08), participants felt significantly more group belonging than in the conditions in which they were part of a different group (M = 3.91, SD = 1.11), $t(213) = 3.43$, $p = .001$, $d = 0.23$. For the Japanese sample, in the vignette conditions in which participants were told they had a relational connection with the victim (M = 2.69, SD = 1.05), participants felt significantly more connected than when there was no relational connection (M = 3.10, SD = 1.14), $t(308) = 5.72$, $p < .001$, $d = 0.32$. In the vignette conditions in which participants were told they belonged to the same group as the victim (M = 3.01, SD = 1.11), participants did not feel significantly more group belonging than in the conditions in which they were part of a different group (M = 2.97, SD = 1.10), $t(308) = -0.58$, $p = .561$. We will come back to this in the Discussion section.

Dependent Variable: Vicarious Anger/Outrage

Vicarious anger/moral outrage, which was the main dependent variable, was assessed with 4 items (αs $\geq .78$ across the four vignettes; e.g., "When I think about what Mr. Parris did,

I feel personally angry") on a 6-point rating scale from 1 = "completely disagree" to 6 = "completely agree" (see Appendix B for all measures). Other emotions measured here were shame, relief, and rumination. These were included as filler items and will not be analyzed here. One item measured participants' vengeful intentions ("If I could, I myself would get back at Mr. Parris for what he did") on the same 6-point rating scale as used for all other items. Vicarious anger/moral outrage was only moderately correlated with vengeful intentions (*r*s = .46 to .49 across the four vignettes; all *p*s < .01). This coincides with the point that these variables have a conceptually different meaning. Thus, the vengeful intention item was analyzed, separately.

Potential Moderators

To measure victim sensitivity (αs > .88), the 10-item scale developed by Schmitt et al. (2005, see also Schmitt et al., 2010) was used. To measure self-representations, the 13-item scale from Brewer and Chen (2007) was used. Individual self-representation was measured with 3 items (αs > .59; e.g., "I am a unique individual"). Relational self-representation was measured with 6 items (αs > .71; e.g., "My happiness depends very much on the happiness of those around me"), and collective self-representation was measured with 4 items (αs > .60; e.g., "The social groups I belong to are an important reflection of who I am"). These items were measured on a 7-point rating scale from 1 = "strongly disagree" to 7 = "strongly agree." For in-group identification we adapted Leach et al.'s (2008) 7-item scale (αs > .91; e.g., "I feel solidarity with my work group"), which was measured on a 7-point rating scale from 1 = "not at all" to 7 = "very much."

Results

Measurement Invariance

To test whether our four-item scale measuring vicarious anger/outrage had comparable properties across the three cultures ("measurement invariance"), we applied confirmatory factor modeling using *Mplus* (Muthén & Muthén, 2007). We followed the step-up approach by Brown (2006) and tested for metric invariance (equal form, equal loadings), which is the required level of measurement invariance for correlation-based analysis. First, we tested measurement models in each country, separately. All models provided a good fit across the four vignettes (ranging from χ^2 = 0.11 to 3.71, *df* = 2, *CFI* = 1, *TLI* = 1, *RMSEA* = .000 to .053, see Table 1). Next, we tested configural invariance, which assumes that the factor structure across all items is identical in all cultures. This proved to show good fit, with the result of the model across the four vignettes, (ranging from χ^2 = 1.72 to 8.20, *df* = 6, *CFI* = 1, *TLI* = 1, *RMSEA* = .000 to .037, see Table 2). In the next step, metric invariance was tested, in which the loadings of the items were constrained to be equal across cultures. These models fit the data well (χ^2 = 17.71 to 31.11, *df* = 12, *CFI* = .988 to 1, *TLI* = .982 to 1, *RMSEA* = .019 to .045, see Table 2). The change in chi-square ($\Delta\chi^2$ = 13.52 to 29.38, *p*s = .000 to .147, Δ*CFI* = .000 to .012) was not significant at $p < .005$ for all but the fourth vignette. Although the change in chi-square was significant here, the change in *CFI* was very close to the critical value of .01 (Cheung & Rensvold, 2002). Furthermore, the various fit indexes suggest that the fit of the model is good in all other cases. We therefore assumed metric invariance for our DV across cultures

Table II-1

Fit of the single country group solutions model testing for vicarious anger/outrage separated by vignette (Vin 1 – Vin 4)

		χ2	df	RMSEA (90% CI)	SRMR	CFI	TLI
USA (n = 214)	Vin 1	1.40	2	.000 (.000- .122)	0.010	1	1
	Vin 2	1.82	2	.000 (.000- .132)	0.011	1	1
	Vin 3	1.63	2	.000 (.000- .128)	0.010	1	1
	Vin 4	0.11	2	.000 (.000- .022)	0.003	1	1
Japan (n = 309)	Vin 1	1.86	2	.000 (.000–.110)	0.009	1	1
	Vin 2	2.96	2	.039 (.000–.127)	0.010	1	1
	Vin 3	3.71	2	.053 (.000–.136)	0.012	1	0.99
	Vin 4	0.82	2	.000 (.000–.087)	0.007	1	1
Germany (n = 274)	Vin 1	3.33	2	.049 (.000- .139)	0.010	1	0.99
	Vin 2	0.28	2	.000 (.000- .062)	0.010	1	1
	Vin 3	2.85	2	.039 (.000- .133)	0.014	1	0.99
	Vin 4	0.79	2	.000 (.000- .091)	0.007	1	1

Table II-2

Measurement invariance testing for vicarious anger/outrage between Germany, USA, and Japan separated by vignettes (Vin 1 – Vin 4)

		χ^2	*df*	χ^2/df	$\Delta\chi^2$	Δdf	*p*	RMSEA (90% CI)	SRMR	CFI	TLI
Equal form	Vin 1	6.59	6					.019 (.000–.084)	0.009	1	1
Equal factor loadings		22.08	12	1.84	15.495	6	0.017	.056 (.014–.093)	0.062	0.990	1
Equal form	Vin 2	5.06	6					.000 (.000–.072)	0.009	1	1
Equal factor loadings		18.58	12	1.55	13.52	6	0.035	.045 (.000–.084)	0.050	1	1
Equal form	Vin 3	8.20	6					.037 (.000–.093)	0.012	1	1
Equal factor loadings		17.71	12	1.48	9.514	6	0.147	.042 (.000–.081)	0.051	1	1
Equal form	Vin 4	1.721	6					.000 (.000–.014)	0.006	1	1
Equal factor loadings		31.11	12	2.59	29.384	6	0.000	.077 (.044–.111)	0.072	0.988	0.982

Hypothesis Tests

Vicarious Anger/Outrage. This DV was submitted to a 2 (relational connection: yes/no; within-subjects) × 2 (group collective connection: yes/no; within-subjects) × 3 (country: Germany, USA, Japan; between-subjects) ANOVA. We found a significant main effect of relational connection, $F(1, 793) = 35.68$, $p < .001$, partial $\eta^2 = .043$. Vicarious anger/outrage was significantly higher when participants knew the victim personally (M = 4.05, SD = 1.20) than when they did not (M = 3.86, SD = 1.24). This effect was qualified by a significant relational connection × culture interaction effect, $F(2, 793) = 4.93$ $p = .007$, partial $\eta^2 = .012$.Breaking down this interaction, the effect of relational connection did not differ between German and US respondents ($p = .085$), but was weaker among Japanese ($t(308) = 1.78$, $p = .076$, $d = 0.10$) compared to US-American and German respondents combined, $t(486) = 6.03$, $p < .001$, $d = 0.27$ (see Figure 1).

The main effect of group collective connection was not significant, $F(1, 793) = 0.74$, $p = .389$, partial $\eta^2 = .001$. Thus, across all cultures, vicarious anger/outrage did not differ as a function of shared group membership with the victim. The group collective connection × culture interaction effect was also not significant, $F(2, 793) = 1.16$, $p = .314$, partial $\eta^2 = .003$. Also, the three-way interaction effect was not significant, $F(2, 793) = .62$, $p = .540$, partial $\eta^2 = .002$.

Vengeful Intentions. Mean values for this single-item measure were analyzed with the same ANOVA model as before. Again, there was a significant main effect of relational connection, $F(1, 792) = 21.73$, $p < .001$, partial $\eta^2 = .027$. Participants reported a higher motivation to vicariously punish the perpetrator when participants knew the victim personally (M = 2.85, SD = 1.51) than when they did not (M = 2.68, SD = 1.47). No other main or interaction effects were significant.

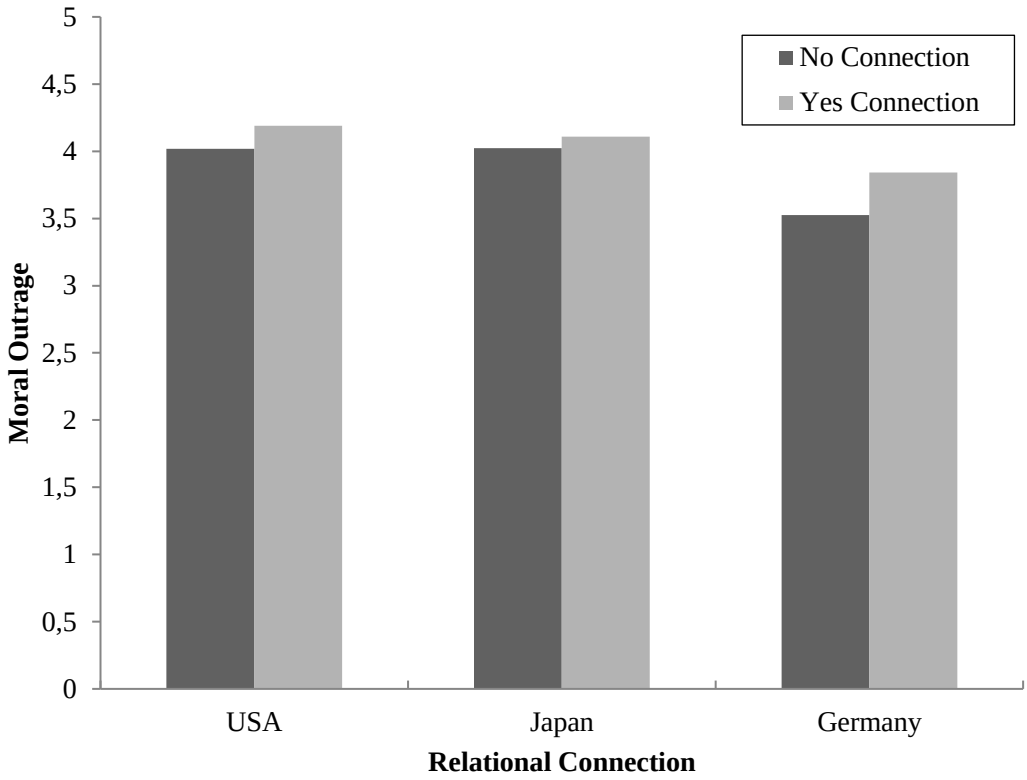

Figure II-1 Relational connection by culture interaction effect. For the participants from Japan, there was less of a difference in moral outrage between those with a relational connection versus without compared to the difference in moral outrage in Germany and USA.

To sum up, we were able to corroborate the notion that relational connections are more relevant for people's vicarious responses to unfairness happening to another person than formal ("group collective") connections. This is in line with Gaertner et al.'s motivational primacy hypothesis: the relational self-representation is more primary to the core of selfhood than the group collective self-representation.

Additional Analyses: Moderation by Individual Differences

Due to the lack of measurement invariance for the four individual differences measures[4], a direct cross-cultural comparison including these measures as moderators of our effects is not appropriate. However, it is still worthwhile to look at within-country effects. First, we looked at vicarious anger/outrage (see Table 3). While the results of the Japan and

[4] The same procedure as above showed non-invariance between cultures for individual self-representation, relational self-representation, collective self-representation, and in-group identification measures. The exception to this was the victim sensitivity measure which was comparable across cultural groups on a level of metric measurement invariance, allowing us to make a meaningful indirect comparison between groups for this measure.

the US samples yielded no significant interactions, the German sample warrants further attention. In the German sample, there is a marginally significant group collective connection × individual self-representation interaction effect (see Figure 2). This is in line with Brewer and Chen's (2007) interpretation of individualism: the higher participants' individual self-representation, the more difference it makes if the victim belongs to the same group as the participant or not. Thus, for Germans as a typical Western ("individualistic") country, individual self-representation is important to collective group membership, but not to relational connections with the victim.

Looking at vengeful intentions, the only individual differences variable that had a significant moderator effect was victim sensitivity: There was a relational connection × victim sensitivity interaction across all cultures, $F(1, 791) = 6.75$, $p = .010$, partial $\eta^2 = .008$ (see Figure 3). The higher the participants' victim sensitivity, the more difference it made if they knew the victim personally or not.

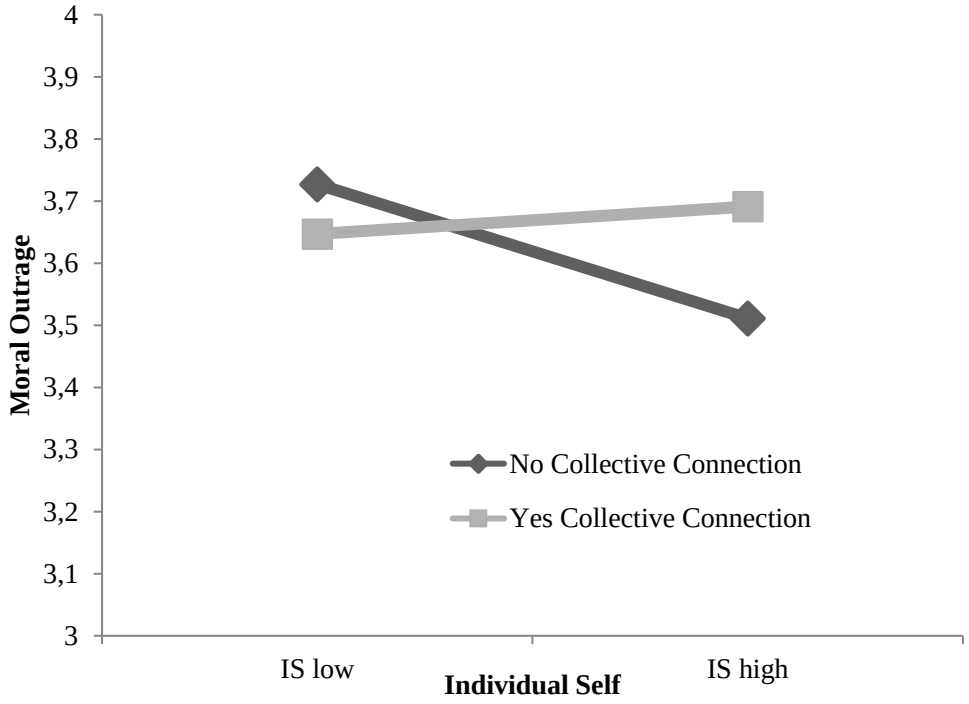

Figure II-2 German sample, significant collective connection by individual self-representation interaction effect. The higher participants' self-reported individual self-representation, the more difference it makes if the group is a collective group connection or not. Individualism is important to collective group categorization in the German sample.

Table II-3

Interaction Effects Between Group Collective and Relational Connection and Victim Sensitivity, Self-Representations, and Identification, Separated by Country

	Germany					Japan					USA				
	df	MS	F	Sig.	Partial η^2	df	MS	F	Sig.	Partial η^2	df	MS	F	Sig.	Partial η^2
Group Collective	1	0.437	0.461	0.50	0.002	1	0.272	0.32	0.58	0.001	1	0.001	0	0.97	< 0.001
* Individual Self	1	3.405	3.593	0.06	0.013	1	1.588	1.85	0.18	0.007	1	1.432	2.11	0.15	0.01
* Relational Self	1	0.065	0.069	0.79	< 0.001	1	2.011	2.34	0.13	0.009	1	0.126	0.19	0.67	0.001
* In-Group Identification	1	3.163	3.338	0.07	0.012	1	0.003	0	0.96	< 0.001	1	0.001	0	0.97	< 0.001
* Victim Sensitivity	1	0.436	0.46	0.50	0.002	1	0.345	0.4	0.53	0.002	1	0.084	0.12	0.73	0.001
Error (Group Collective)	268	0.947				254	0.861				209	0.679			
Relational	1	19.47	22.63	0.00	0.078	1	1.472	2.08	0.15	0.008	1	5.93	7.03	0.01	0.033
* Individual Self	1	0.27	0.314	0.58	0.001	1	0.091	0.13	0.72	0.001	1	1.523	1.81	0.18	0.009
* Relational Self	1	0.635	0.739	0.39	0.003	1	0.741	1.05	0.31	0.004	1	0.13	0.15	0.70	0.001
* In-Group Identification	1	0.051	0.059	0.81	< 0.001	1	1.294	1.83	0.18	0.007	1	2.698	3.2	0.08	0.015
* Victim Sensitivity	1	2.244	2.609	0.11	0.01	1	0.021	0.03	0.86	< 0.001	1	0.584	0.69	0.41	0.003
Error (Relational)	268	0.86				254	0.707				209	0.844			
Group Collective * Relational	1	0.001	0.001	0.97	< 0.001	1	0.267	0.38	0.54	0.001	1	0.49	0.73	0.40	0.003
Error (Collective * Relational)	268	0.899				254	0.709				209	0.673			

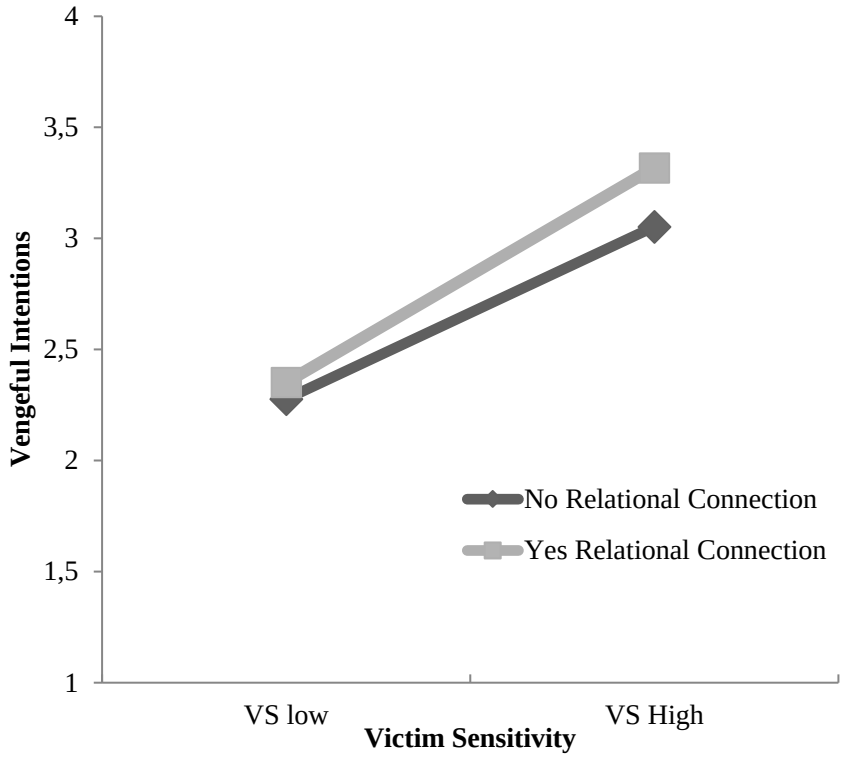

Figure II-3 Relational connection x victim sensitivity interaction across all cultures. For participants high in victim sensitivity, there was a greater difference between vengeful intentions for the condition with a relational connection compared to those without.

Discussion

When discussing the psychological meaning of the "self" there is often, but not always, an overlap in different self-representations. Our friends, family and co-workers are conceivably in our same social collective, be that university, church group, or company. However, even with some convergence, the three fundamental selves are conceptually distinguishable. If we wish to definitively rank the relative importance or to identify cross-cultural differences, it is important to experientially parse out the distinct definitions and reevaluate the relative primacy. That is precisely what this study has done.

Brewer and Chen (2007) drew attention to the importance of the definition of the group in understanding cross-cultural differences in self-representations. Beyond the individual self, which is straightforward, they argued that the collective self needs to be further distinguished into relational self-representations (based on personal relations with

others) and group collective self-representations (based on shared group memberships). The present study tested whether relational self-representations are more relevant than group collective self-representations (as argued by the motivational primacy hypothesis; cf. Gaertner et al., 2012), and whether this holds across different cultures. Most importantly, our study is, to the best of our knowledge, the first to test this hypothesis with a stringent design – that is, by manipulating relational and group collective connections orthogonally. Under these conditions, we found further evidence for the hierarchy of the relational self over the group collective self. Participants across three cultures were angrier when a relational other was harmed than when a collective group member was harmed.

There was minimal support for a cultural context explanation given that the only significant interaction effect including culture was that Japanese respondents tended to be less sensitive to the "relational connection" variation than respondents from the other two countries. This is an interesting finding because we would have expected Japanese respondents to be *more* sensitive to relational connections (rather than to group collective connections, cf. Yuki et al., 2005). We can only speculate about the reason for this. For instance, it could be that for Japanese participants, "work department" might have been too small as to be a meaningful social unit. Support for this can been seen in the manipulation check where participants from Japan did not feel more belonging when they were in the same group compared to a different group. For Japanese participants, it is possible that membership in the same work department was not distinctive enough to parse out a relational connection; those in a different department are still part of the same company and therefore plausibly still in the same social network. Therefore, we might have observed less differentiation across all conditions in the Japanese sample. Further research should look to make the collective group distinction more concrete, especially in an Asian context. Although this interaction effect hints at future exploration of a cultural element in the relative importance of relational connections, overall the evidence for a pan-cultural hierarchy is convincing.

Limitations

One aspect that may be considered a limitation of the present study was the discrepancies between data collection in three samples. More precisely, in the German sample participants responded to a call for volunteers with a lottery, while the participants in the other two countries were recruited through a third-party platform and all received a small payment via that platform. Part of this is due to the challenges of collecting culturally diverse non-student sample data. However, despite this limitation, we found strong evidence for the relational self over the group collective self in each setting. Therefore, the main finding is persuasive over any alternative explanations related to sample recruitment.

Another concern was the lack of a comparable individual difference measure. Although different nations are categorized by cultural differences, individuals will adopt these cultural norms to different degrees. Therefore, it is important to measure the individual endorsement of cultural ideals. In this aspect our study was limited. A more reliable self-representation measure could better detect individual difference in self-representation, which in turn could support the hypothesis of relative importance of the relational self or the group collective self in people with different cultural orientations, even if the three tiered hierarchy remains intact.

Finally, although the vignette presented different workplace conflicts, they were of the same format and designed specifically for this task; therefore, our vignettes might be considered limited in their generalizability. However, this was not unintended. Rather than using an established paradigm, this study was designed to be additive to past research and present previously unexamined circumstances for which to observe the motivational hierarchy of self-representations.

Conclusion

The inclusion of others in one's own self-concept can be observed across cultural contexts. This can be reflected in the emotional reaction to observed injustice. This paper investigated the relative importance of self-representation beyond the individual self that includes relational others and those with whom one shares a common group membership. Across three cultures, the extent to which participants were relationally connected to a victim (i.e., knew the victim personally) was a better predictor of their experiences of vicarious anger and moral outrage as well as their inclination to vicariously punish the perpetrator. This is the case in both Western and Eastern cultures: the relational self holds primacy over the group collective self-representation.

References

Brewer, M. B., & Chen, Y.-R. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological Review, 114*, 133–151.

Brewer, M. B., & Gardner, W. (1996). Who is this "We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology*, *71*, 83–93.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, *1*, 185-216.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.

Chen, Y. R., Chen, X. P., & Portnoy, R. (2009). To whom do positive norm and negative norm of reciprocity apply? Effects of inequitable offer, relationship, and relational-self orientation. *Journal of Experimental Social Psychology*, *45*, 24–34.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measure-ment invariance. *Structural Equation Modeling*, 9, 233–255.

Cross, S. E., & Madson, L. (1997). Models of the self: self-construals and gender. *Psychological bulletin*, *122*, 5-37.

Gaertner, L., Sedikides, C., & Graetz, K. (1999). In search of self-definition: Motivational primacy of the individual self, motivational primacy of the collective self, or contextual primacy? *Journal of Personality and Social Psychology,* 76, 5–18.

Gaertner, L., Sedikides, C., Luke, M., O'Mara, E. M., Iuzzini, J., Jackson, L. E., ... & Wu, Q. (2012). A motivational hierarchy within: Primacy of the individual self, relational self, or collective self?. *Journal of Experimental Social Psychology, 48*, 997-1013.

Gaertner, L., Sedikides, C., Vevea, J., & Iuzzini, J. (2002). The "I", the "We", and the "When": A meta-analysis of motivational primacy in self-definition. *Journal of Personality and Social Psychology, 83,* 574–591.

Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J., & Baer, A. (2005). Asymmetrical Effects of Justice Sensitivity Perspectives on Prosocial and Antisocial Behavior. *Social Justice Research*, *18*, 183–201.

Lane Beckes, James A. Coan, Karen Hasselmo; Familiarity promotes the blurring of self and other in the neural representation of threat, *Social Cognitive and Affective Neuroscience*, Volume 8, Issue 6, 1 August 2013, Pages 670–677

Leach, C. W., van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., … Spears, R. (2008). Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, *95*, 144–165.

Lee, S., Adair, W. L., Mannix, E. A., & Kim, J. (2012). The relational versus collective "We" and intergroup allocation: The role of nested group categorization. *Journal of Experimental Social Psychology*, *48*, 1132–1138.

Leung, A. K. Y., & Cohen, D. (2011). Within-and between-culture variation: individual differences and the cultural logics of honor, face, and dignity cultures. *Journal of personality and social psychology*, *100*, 507-526.

Loewenstein, G., & Small, D. A. (2007). The Scarecrow and the Tin Man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, 11(2), 112.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98*, 224-253.

McCrea, S. M., Wieber, F., & Myers, A. L. (2012). Construal level mind-sets moderate self- and social stereotyping. *Journal of personality and social psychology*, 102(1), 51.

Muthén, L. K., & Muthén, B. O. (2007). Mplus. Statistical analysis with latent variables. Version, 3.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological*

Bulletin, *128*, 3–72.

Realo, A., & Allik, J. (2009). On the relationship between social capital and individualism–collectivism. *Social and Personality Psychology Compass, 3*, 871–886.

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity: Assessment and location in the personality space. *European Journal of Psychological Assessment*, *21*, 202–211.

Sussenbach, P., & Gollwitzer, M. (2014). Us(ed): The role of victim sensitivity in potentially exploitative intergroup relationships. *Group Processes & Intergroup Relations*, *18*, 241–255.

Triandis, H. C. (1995). *Individualism & collectivism*. Westview press.

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological review*, *96*, 506-520.

Yuki, M. (2003). Intergroup Comparison versus Intragroup Relationships: A Cross-Cultural Examination of Social Identity Theory in North American and East Asian Cultural Contexts. *Social Psychology Quarterly*, *66*, 166-183.

Yuki, M., Maddux, W. W., Brewer, M. B., & Takemura, K. (2005). Cross-cultural differences in relationship- and group-based trust. *Personality and Social Psychology Bulletin*, *31*, 48–62.

Appendix A

Full text of all the vignettes

1) "A member from another team within your large company (let's call him Mr. Smith) approaches someone from **your department** that *you don't know personally* (let's call him Jack) and asks to switch shifts with him because he needs to see a doctor. Jack agrees and takes Mr. Smith's early-morning shift. Later you hear that Mr. Smith actually never saw a doctor (and never intended to do so); he just wanted to sleep in that day because he had been partying the night before."

2) "Imagine you work for a large company. Your boss (let's call him Mr. Parris) calls in someone from **your department** that *you know personally* (let's call him John) to his office to discuss their time sheet. There were days in the past month that John worked from home. However, Mr. Parris argues that John didn't have permission to do so, so he discounts some of John's work hours. Mr. Parris was aware that many other colleagues also work from home in the past month but they were not reprimanded."

3) "Imagine you work in a large company. Employees share a person who helps with administrative tasks (let's call this person Mrs. Park). Mrs. Park is most unhelpful to someone from **another department** that *you don't know personally* (let's call this person Jessica) because she dislikes Jessica personally. Every time Jessica asks Mrs. Park to complete a task she says she is busy and will attend to it later. However, when colleagues ask for something Pearl attends to it immediately. This affects the quality and efficacy of Jessica's work."

4) "Imagine you work in a mid-size company. Your boss asks someone from **another department** that *you know personally* (let's call this person Jane) to work with another employee (let's call her Mrs. Johnson) on an important report that will raise Jane's profile at the company. The two spend weeks going back and forth collaborating on the report. Although they worked well together on this report, later Jane learns that Mrs. Johnson is spreading rumors about her and her work ethic, saying she is lazy and combative in hopes Patty will be asked to work on the next report without Jane, further improving Mrs. Johnson's own reputation within the company."

Appendix B
Measures

1. "When I think about what Mr. Smith did, I feel personally angry" *
2. "I have no strong negative emotions in reaction to this situation" (r) *
3. "I feel contempt with Mr. Smith's behavior" *
4. "Personally, I feel outraged about what Mr. Smith did" *
5. "I am personally relieved I am not in Jack's position"
6. "I feel personally ashamed for Mr. Smith"
7. "I will ruminate about what happened to Jack for some time now"
8. "If I could, I myself would get back at Mr. Smith for what he did" †
9. "What Mr. Smith did to Jack feels as if he did it to me"
10. "I would not think twice about what Mr. Smith did" (r)

Completely Disagree 123456 Completely Agree

* Items for vicarious anger/moral outrage

† Item for vengeful intentions

III. Chapter 3: Expansion of the Self in the Reaction to Injustice of Others: Categorical versus Relational Self

Zoe Magraw-Mickelson, Leonie Schmitt, Paulina Stocker and Mario Gollwitzer

Abstract

When do people feel the most concern for the injustice towards others? Based on findings from the self-construal literature, we hypothesize that people, as observers, are most concerned about third-party injustice when there is a fit between their self-construal and the kind of relationship that they have with the victim -- in other words, observers will be most outraged when primed with a relational self-construal, learn about a relational group member being victimized, and when observers primed with an individual self-construal learn about a group member with a shared category being victimized. In an experimental lab study with Chinese university students (N=141), we took a novel approach to manipulating relational groups in addition to categorical groups, using a minimal group paradigm, as well as manipulating self-construal using a priming task. Although we did not find evidence for our main hypothesis when using the primed self-construal, we found indirect support looking at the individuals' measured self-construal. Specifically, interaction between participants' interdependent self-construal and type of shared group predicted emotional outrage in observed victimization of a third party other.

Introduction

In unfair situations people are likely to feel concerned about the injustice to oneself, but when do people feel the most concern for the injustice towards others? To answer this question we first need to focus on who is implied in 'others' and how that relates to personality characteristics. People will have more concern when the victim has some connection to the self, compared to no connection, but what type of connections are important? Based on findings from the self-construal literature, people are most concerned about third-party injustice when there is a fit between one's self-construal and the kind of relationship that the observer has with the victim. The present study will look at how the degree of concern for third-party injustice is influenced by group membership, either categorical or relational, in conjunction with self-construal, either independent or relational.

Differences in self-construal signify aspects of one's self definition, specifically the degree to which people think of themselves in terms of others. These dispositional traits are categorized as independent and interdependent self-construal style[5] (Cross, Bacon, & Morris, 2000). Past research has suggested that individuals with a more interdependent self-construal include others in the self, while individuals with a more independent self-construal are less likely to include others in the self (Gardner, Gabriel, & Hochschild, 2002). However, this is a simplistic explanation, especially in that it does not extrapolate on who specifically is important for those with an interdependent self-construal. Research also suggests that groups, and group bonds are also important to those thought of as individualistic (Social Identity Theory: Tajfel & Turner, 1979). Therefore, we need to look at the types of group and personality factors in conjunction.

[5] Although independent and interdependent self-construal are often theoretically understood as orthogonal constructs (e.g. Markus & Kitayama, 1991; Oyserman, 1993; Singelis, 1994; Triandis, 1995), the evidence is mixed. Meta-analysis has shown that the dimensionality of the constructs depend on a number of factors, such as the sample characteristics, the measure being used, and the level of analysis (Taras et al. 2014). Nonetheless, as Taras and colleague (2014) point out, even when conceptualized as independent constructs the definitions are composed of highlighting the opposite features. We will also refer to the constructs as opposites but consistently test both constructs independently.

This study will look at self-construal, both measured and manipulated, and group type using minimal group paradigms in situations of third-party injustice. Concern for third-party injustice is the extent to which the unfair-fair difference differs between oneself and another person being the victim. This captures concern for the injustice of others as it is contrasted to the self, therefore providing insight into who are the relevant 'others' when talking about the self, group membership and injustice.

Self-construal

Self-construal describes ways in which the self is defined; more specifically the self-construal framework describes independent and interdependent self-construal influences on cognition, emotion and behavior. Markus and Kitayama (1991) popularized the two complimenting constructs now widely researched in cross-cultural studies. As defined in the literature: interdependent self-construal represents individuals who define themselves in terms of relationships and integrate others into their self-concept, while individuals with a more independent self-construal place less emphasis on social embeddedness with the group (Brewer & Gardner, 1996; Markus & Kitayama, 1991). Each self likely coexists in an individual; however, one self-construal is more likely to be more dominant then the other (Gaertner et al., 2012; Hannover & Kühnen, 2004; Singelis et al., 1999). Although self-construal varies between cultures, the relative endorsement of dispositional self-construal also varies within cultures (Cross et al., 2000; Singelis, Bond, Sharkey, & Lai, 1999; Singelis, 1994).

Key to these definitions is the focus on the independence or the integration of the self with others or the group, however, it is still not conceptually clear who is meant by 'others' or 'group'. Brewer and Chen, (2007) convincingly argue that the literature is vague on this point and call for differentiation in types of in-group relationships. Inconsistent findings in the literature with regard to the construct (i.e., Oyserman, Coon, & Kemmelmeier, 2002), also gives weight to the notion that there is a necessity to differentiate the meaning of others and

the group. Specifying what is meant when defining interdependence or collectivism as the presence or absence of a connection to the group, will clarify our understanding of self-construal.

In more detail, it has been proposed that collectivism should be conceptually divided into the relational self and categorical self, to highlight different types of groups or others with whom one can be connected. In terms of self-construal, relational self is specified as the tendency for the self to be defined by the connection to a group due to a personal relation (such as a close friend or family member) while categorical-self is specified as the tendency for the self to be defined as the connection to the larger collective group with whom one shares a category social (such as, someone from the same university or geographical region; cf. Brewer & Chen, 2007; Brewer & Gardner, 1996; Yuki, 2003). Therefore, self-construal can be defined as "individual self", "relational self", and "categorical self". The relational self is more conceptually in line with the original concept of the interdependent self but highlights the importance of those with whom one has a personal connection to one's self-concepts. Categorical self, on the other hand, was theorized by Brewer and Chen (2007) to be relevant for those traditionally described as independent. Groups can also be defined as "relational", meaning groups of people who have an interconnected relationship, and as "categorical", meaning groups of people who share the same label. With these distinctions we can better predict the interaction of personality characteristics and group type.

Past research has shown an overall motivational hierarchy of the self across cultures (Gaertner et al., 1999, 2002, 2012; Magraw-Mickelson & Gollwitzer; 2018). That is, overall the individual self is more important than the relational self, and the relational self is more important than the collective self. That is, across culture when the different aspects of the self are in conflict people are most concern for the individual self and least concerns for the collective self. However, the similarity across cultures in the importance of the three self-representations does not answer the question about the interaction between the relative

importance of the self and group type. The current study will ask this question by looking at self-construal and group type within a Chinese sample.

Furthermore, although research has found evidence of the differentiation between aspects of collective self, there has not been many explicit tests of the relationship between individual self-construal and categorical groups. It is theorized that even cultures that are more individualistic also care about groups, namely those which they share a group membership (Brewer & Chen, 2007). Magraw-Mickelson and Gollwitzer (2018) found that within the German sample, participants high in independent self showed more outrage when a victim was from their collective group compared to not part of the collective group; however, this was not found in the American or Japanese samples. The present study will follow up on these results by directly testing the relationship between individual self and categorical groups, expecting their relationship to be similarly matched as relational self and relational groups.

Therefore, when we look at for whom one is concerned in situations of injustice, there will be most concern when there is an overlap between the self and the other, that is, when there is a match between the type of group which is important to the self and the group at hand. We will find less outrage if there is a mismatch.

Minimal – Groups

In order to investigate group types we will look to the minimal group literature. Minimal group paradigms are often used in social identity theory research to investigate in-group favoritism and are defined as "depersonalized social categories based on arbitrary category distinctions between in-group and out-group" (Brewer & Chen, 2007, p. 147). It has been found in Western countries that artificially created groups can induce in-group favoritism based solely on categorization, demonstrating the dominance of in-group bias in intergroup evaluations and allocation behavior (Perreault & Bourhis, 1998; Platow, McClintock, & Liebrand, 1990).

Some research has considered if the relationship between in-group identification and in-group bias depends on personality factors such as self-construal. In studies that examined whether the identification and in-group bias relationship was moderated by collectivism, Capozza, Voci, and Licciardello, (2000) found, counter to predictions based on the importance of groups for collectivist, that the bias was stronger and more coherent for individualists. These unexpected results were explained as relating to differences in self-esteem motivation, that is, in-group identification and enhancement contribute to self-esteem which is more important to those with an independent self-construal, however this was not directly tested. Other studies have found that in Japan, a typical collectivist culture, categorization alone did not induce in-group favoritism, rather participants only showed in-group favoritism in the experimental condition in which there was an expectation of generalized reciprocity between in-group members (Yamagishi, Jin, & Miller, 1998). Further research supported this finding that East Asians do not readily engage in outgroup discrimination in minimal group settings, when discrimination does not indirectly benefit the self (Brewer & Chen, 2007; Yamagishi, Jin, & Kiyonari 1999).

However, it is clear that in past research minimal groups relied on only one form of group membership, that is, the *categorical* minimal groups. In these studies researchers assigned group membership based on an arbitrary label and found results such as positive in-group bias and empathy, as well as out-group discrimination (Montalan, Lelard, Godefroy, & Mouras, 2012; Reynolds, Turner, Haslam, Ryan, Bizumic, & Subasic, 2007; Rubini, Moscatelli, & Palmonari, 2007). In-group favoritism and bias can also occur in groups that are relationally formed. It is not unreasonable to assume that people will also show favoritism toward those with whom they know even in the absence of a categorical label. Therefore, in addition to traditionally formed categorical minimal groups, we also form relational minimal groups by inducing a personal relationship between strangers in an experimental setting.

In addition, minimal groups are advantageous in this design because in real life there is often an overlap between group types, that is someone who belong to the same categorical group is also someone one knows personally. Minimal group paradigms minimize the noise from real group's membership categories and intersecting relationships.

Priming Self-Construal

Self-construal is an aspect of personality that has been shown to differ between cultures; however, there is also within-country variation. Furthermore, self-construal can also be experimentally manipulated by making the self-knowledge associated with one or another self-construal temporarily accessible. It is theorized that people from individualists and collectivists countries respond to personal and group treatment according to mindsets that are the cultural "default". Priming self-construal allows researchers to move away from a reliance on national culture as a proxy for self-construal (Cross et al, 2011). Studies have demonstrated that situational cues can lead to "frame switching" between the different self-construal in bicultural participants (Hong, Morris, Chiu, & Benet-Martinez, 2000). Thus, priming can be used to experimentally examine the effects of individual self-construal, relational self-construal, and categorical self-construal on behavior (Cross, Hardin, & Gercek-Swing, 2011). This idea of priming culture such as collectivism and individualism has been explored extensively (for a review see Oyserman & Lee, 2008). Counter-cultural priming paradigms in cross-cultural research not only show why cultures may differ, they also provide a high degree of internal validity (Van den Bos, Brockner, van den Oudenalder, Kamble, & Nasabi, 2013). One of the most common priming techniques is the pronoun circling task. The concept is that pronouns, such as "we" and "us", carry emotional significance and meaning, and that these characteristics can automatically and unconsciously shape self-construal (Perdue, Dovidio, Gurtman, & Tyler, 1990).

Priming is a useful procedure to delineate the causal relationship of cultural differences and psychological dimensions (Van den Bos et al., 2013; Van den Bos, Van

Veldhuizen, & Au, 2014). However, there are a few limitations to current priming strategies, such as the ambiguous nature of current priming techniques, especially in regard to the differentiation between priming relational-self and categorical-self which have been attempted much less extensively than independent-interdependent self-construal. Most of the manipulations used in past studies make salient relational aspects of the self (Gardner , Shira, & Lee, 1999) and the way the self is similar to others (Trafimow et al., 1991), excluding obligation or duty to the categorical in-group.

Although some research indicates the effects of priming tasks are similar for East Asians, Europeans, and Americans (Oyserman & Lee, 2008), relatively few studies have used the priming techniques with non-Western samples. In addition, when priming is investigated in non-Western countries it is usually a comparison between a Western sample and a Japanese or Korean, or Hong Kong Chinese samples. Therefore, more studies are needed to assess the effectiveness of the priming tasks for other non-Western countries and diverse Asian samples (Cross et al, 2011).

Justice sensitivity

Finally, when investigating perceptions of injustice it is important to include other theoretically-related facets of personality. Justice sensitivity describes differences in the degree to which individuals are concerned with injustice (Gollwitzer, Schmitt, Schalke, Maes, & Baer, 2005). It has been found to be a highly stable and consistent personality trait. A heightened sensitivity towards injustice can be from different perspectives, such as from the perspective of an observer, victim, beneficiary or perpetrator. Observer sensitivity, which is the sensitivity to injustice from the observer's perspective, is expected to predict the degree of concern for the injustice of others regardless of the relationship to the victim. Therefore, we would expect observer sensitivity to have a direct relationship to emotional outrage at witnessing others' disadvantage. Victim sensitivity on the other hand, which is the extent to which people react emotionally when confronted with injustice to their own disadvantage and

the degree to which one is susceptible to feeling like a victim may also be relevant to the question at hand in some conditions (Süssenbach & Gollwitzer, 2014).

It has been shown that people high in victim sensitivity are sensitive not only to their own injustice but also to that of their group members, both for categorical groups (Süssenbach & Gollwitzer, 2014) and relational groups (Magraw-Mickelson & Gollwitzer, 2018). The concepts of self-construal in conjunction with victim sensitivity means that, especially for people high in victim sensitivity, if a self-construal is relevant to an individual, others that belong to that category will be included in the self which will result in increased emotional outrage when that person is the target of injustice, compared to victims who are not part of their expanded self.

Present research

The present research will integrate the above concepts to investigate the question of when people are most concerned for the injustice of others. The hypothesis here is that people are most concerned about third-party injustice when there is a fit between self-construal and the kind of relationship that observers has with the victim. This is especially true for those high in victim sensitivity. Self-construal is both measured via self-construal scales and experimentally manipulated via a priming procedure. The priming manipulation consists of three conditions (individual, relational, and neutral control). The control condition is included to establish a baseline. Relationship to the victim is defined as whether the victim and the observer share a categorical group membership or a relationally-connected group, which is defined by personal connections. A minimal group paradigm will be used to make groups which are either categorical or relationally defined.

Our primary research question is whether self-construal and the kind of relationship to the victim interact with each other, such that concerns about another person being treated unfairly is highest (a) in the individual self / categorical group and (b) in the relational self / relational group conditions, That is:

Hypothesis 1: When primed with an individual self, people will experience stronger justice-related emotions ("anger" and "moral outrage") and show more punitive behavior ("punishment" at a cost) when a categorical (vs. relational) group member is harmed compared to the control or relational prime conditions.

Hypothesis 2: When primed with a relational self, people will experience stronger justice-related emotions ("anger" and "moral outrage") and show more punitive behavior ("punishment" at a cost) when a relational (vs. categorical) group member is harmed compared to the control or individual prime conditions.

In addition, we explored a secondary research question: To what extent does victim sensitivity moderates the main effect.[6]

Methods

Participants

Prior to data collection, we conducted a power analysis estimation using G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007). This analysis suggested a sample size between 116 ($\eta^2 = .085$; $\alpha=.05$; $1- \beta =.90$) and 180 ($\eta^2=.055$; $\alpha=.05$; $1- \beta =.90$) participants. Taking possible dropouts and mistakes during data collection into account, 176 participants were recruited from a large Chinese university. Thirty-eight participants were excluded from analysis using previously establish criteria: not of Chinese nationality (4), missing data (4), failed attention check[7] (30). Five additional participants were excluded because of an experimenter's errors (e.g., the experimenter did not give the participant the correct materials). The remaining sample consisted of 141 participants (57% females, 43% males). Participants were between 17 and 28 years old (M=20.70, SD=2.30). Most of the students were undergraduates (82%; 13% graduates, 4% post graduates, 1% High school student).

[6] See pre-registration of design materials and hypotheses at https://osf.io/9qgeh/?view_only=2494c503827a4b1b8cfd1a8ffef3be2b

[7] The survey portion of the materials contained two attention check items (e.g. "This is an attention-check. Please tick "exactly.").

Participants were recruited through the university's online platform, on campus and in classes. The advertisement introduced a study that explored the influence of personality traits in groups and behavior and stated the possibility to earn up to 16 yuan for participation.

Experimental procedure

In total, seven experimenters conducted 50 experimental sessions of 3-5 participants[8] that were randomly assigned to one of the six conditions[9]. Experimenters followed a standardized script, each experimental session lasted 30 minutes and was conducted as described in the following.

To begin, participants signed the informed consent, which assured that participants knew about the frame of the study as well as their personal rights within the study. This was followed by three pages of questionnaires containing the variables: victim sensitivity, observer sensitivity, individual self-construal, interdependent self-construal and identity as student group scale (see below for details) and demographic information.

After the completion of the questionnaires, participants then learned about an economic game that they would play in two rounds with an unknown counterpart and were given an answer sheet to record how they would play in the games. As part of the cover story, participants were told that, depending on the decisions made in these two rounds, they could earn additional money. Participants took part in a modified trust game (see e.g. Berg, Dickhaut, & McCabe, 1995) which is played as follows: in the first round of the game, participants were assigned the role of the sender. As such, participants received 4 tokens (1 token = 2 yuan) some of which they could send to an unknown receiver. Any amount of tokens they sent would be tripled (e.g. sent 2 tokens, receiver would get 6 tokens). Later, in

[8] In cases where only 2 participants were scheduled for an experimental session a confederate acted as a third participant.

[9] Conditions here refer to the four experimental conditions (individual SC & categorical GT, relational SC & categorical GT; individual SC & relational GT, relational SC & relational GT) as well as control conditions (neutral SC & categorical GT; neutral SC & relational GT).

the second round of the game the receiver could decide whether to keep the tokens or share them with the sender.

Minimal group paradigm. After collecting the game responses, the experimenter told participants that "I will now leave to get the results of several games and information relating to the measures you have completed. While I am gone, I want you to get to know the other participants in the room. Talk to all of them and find out whether you have a friend in common or how else you might know each other." The experimenter asked them to sit together at one table for that purpose and left the room for 7 minutes. Conversations between participants during that time served to create attachment among the participants, which is a fundamental feature of a relational group. After that time of getting to know each other, the experimenter returned to the room, and the participants were informed that based on earlier completed measures, they were assigned to either categorical group A or group B, in a minimal group paradigm (see e.g., Reynolds et al., 2007, in fact, all participants were assigned to group A). To intensify the categorical group information, participants had to write their categorical group (e.g. group A) on each of the following questionnaire pages. In addition the experimenter said, "Please note your group. In this experimental session, participants in the room are split between the two groups."

Priming Self-Concept. To prime self-construal, we used the pronoun circling task (Gardner et al. 1999; Kashima, Hardie, Wakimoto, & Kashima, 2011; Kühnen, Hannover, & Schubert, 2001). In this procedure, participants were asked to read a paragraph about a day in the city (see Brewer & Gardner, 1996). Depending on the condition, the paragraph contained either 18 "individual pronouns", 18 "relational pronouns", or 17 "vehicle words" (taxis, trucks, cars, vehicles) that participants had to circle when they read the paragraph for the second time. The description of the manipulation conditions are as follows (see Appendix A for full text):

1. Individual self: The individual prime condition was told from the first-person singular perspective, using first-person singular pronouns only (I, me, my).

2. Relational self: The relational prime condition was told the story from the perspective of "my partner and I," using first-person plural pronouns only (we, us, our).

3. No prime neutral condition: In the no prime condition, the story was told from a neutral perspective.

Trust Game, round two group manipulation. Immediately after the priming of the self-construal, the second round of the trust game started. In this round participants saw the results of 5 games shown in Table 1. In each game, the sender gave 4 tokens, meaning the receiver had 12 tokens to disperse.

Table III-1

The results of the trust game for the 5 games

	Game 1	Game 2	Game 3	Game 4	Game 5
Sender	In-group Member 1	In-group Member 2	Participant	Participant	Out-group Member 6
Receiver	Out-group Member 1	Out-group Member 2	Out-group Member 3	Out-group Member 4	Out-group Member 5
Tokens sent back to Sender	2	6	2	6	4
Tokens kept by Receiver	10	6	10	6	8
Distributional fairness	Unfair	Fair	Unfair	Fair	Unfair

Depending on the conditions, players affected by the game results were either another person from their same group A (categorical group) or another person in the same session (relational group). For example, the result of game 1 for a participant in the categorical group condition read as follows:

> "These are results of a game between a different Group A member (Sender) and a different Group B member (Receiver). A Group A member as the Sender gave all 4 of the tokens, so the Group B member as the Receiver received 12 tokens (4 tokens tripled by the experimenter). The Group B Receiver decided to return 2 tokens to the

Sender. Result: This Group A player has 2 tokens and the Group B member has 10 tokens."

After each of the five game results, participants filled out measures of anger and moral outrage. Participants also had the opportunity to punish the receiver at their own cost, by losing 3 tokens for each token they would take away from the receiver. In the final portion of the study participants completed a manipulation check for the group manipulation and for the priming manipulation, as well as an inclusion of the other in the self (IOS) scale for both manipulated groups (see below).

Debriefing and reward. At the end of the experiment, participants were revealed the real purpose of the lab experiment and received money for participation.

Materials

Translation process. Original English instructions and measures served as a basis to obtain valid Chinese equivalents[10].

Dependent variables. Anger: Anger was measured via self-report using an adaptation of the State-Anger subscale Ärgergefühl of the State-Trait-Ärgerausdrucks-Inventar-2 [State-Trait-Anger-Expression-Inventory-2] (STAXI-2, Rohrmann et al., 2013). The five items (e.g., "I am mad.") were rated on a 6-point Likert-scale ranging from 0 = "not at all" to 5 = "exactly" $\alpha = .87$.

Moral outrage: Moral outrage was measured via self-report using two items adapted from Gollwitzer, Braun, Funk, and Süssenbach (2016) and three items developed for this study. The five items (e.g. "I am indignant.") were also rated on a 6-point Likert-scale ranging from 0 = "not at all" to 5 = "exactly". $\alpha = .91$.

[10] We follow the basic back translation procedure outline in Brislin, (1970). People fluent in both languages each engaged in one of the different steps of the following translation process: first the original questionnaires were translated from English to Chinese, and then back translated from Chinese to English. In the next step, discrepancies between the original English version and the first English back translation were highlighted. If those discrepancies were due to mistakes in the Chinese translation, the respective sentences were translated (English - Chinese) and retranslated (Chinese - English) until both the original English version and the Chinese version were found to be equivalent. Finally, the Chinese questionnaire was pre-tested which resulted in minor final improvements of the instructions and measures.

Punitive behavior: Punitive behavior was measured by asking participants if and (if yes) with how many tokens they wished to punish the receiver considering a punishment rate of 3:1 at their own cost. The maximum possible punishment was indicated in parenthesis, as the game numbers differed between games (10 vs. 6 vs. 8 tokens). It also served to compute the proportional punishment (actual punishment/possible punishment) for each game.

Hypotheses.

We expect a two-way interaction effect (priming: individual, relational, neutral) x relationship/group type (categorical, relational) on the DV's (second-order difference score) regarding emotional and behavioral reactions;

H1: Participants primed with an individual-self will experience stronger justice-related emotions ("anger" and "moral outrage") when a categorical (vs. relational) group member is harmed compared to the control or relational prime conditions.

H2: Participants primed with a relational-self experience stronger justice-related emotions ("anger" and "moral outrage") when a relational (vs. categorical) group member is harmed compared to the control or individual prime conditions.

Supplementary variables.

Potential moderator. The personality trait victim sensitivity was assessed using the Chinese version of the subscale Victim Sensitivity of the Justice Sensitivity Inventory (Schmitt, Gollwitzer, Maes, & Arbach, 2005). Participants rated the ten items (e.g., "It makes me angry when others receive a reward that I have earned.") on a 6-point Likert-scale from 0 = "not at all" to 6 = "exactly". $\alpha = .85$.

Potential covariates. The personality trait observer sensitivity (10-items subscale Observer Sensitivity of the Justice Sensitivity Inventory; Schmitt et al., 2005) $\alpha = .85$, independent self-construal style and interdependent self-construal style (5-items subscale Autonomous Orientation subscale and 4-items subscale Social Orientation of the Relationalitäts-Kontextabhängigkeits-Skala [Relationality-Dependency-Scale]; Gollwitzer,

Schmidthals, & Pöhlmann, 2006), α = .78 and α = .60 respectively, and participants' identity with students as a group (adapted 6 items scale Group identity; Leach et al., 2008) were all rated on a 6-point Likert-scale from 0 "not at all" to 6 = "exactly" α = .85.

Priming manipulation check. To assess participants' activated type of self-construal after the pronoun circling task, participants completed an adapted version of Kuhn & McPartland's (1954) Twenty-Statements-Test, by writing 10 statements in response to the question "Who am I?" (as in, e.g., Kashima et al., 2011).

Check of following manipulation instructions: Two items asked participants which categorical group they were in and whether they talked to participants from their experimental session when instructed to do so.

Potential mediator. To see whether participants extended their self to include others in their self as a consequence of the manipulated group identity, participants answered two items adapted from Aron, Aron, Tudor & Nelson's (1991) Inclusion of Other in the Self scale which asked about the overlap between the self and the "people with you in this session" and "other group A members".

Results

Manipulation check

For the "Who am I?" task, we calculated the proportional number of individual, relational, and categorical statements (see e.g. Kashima et al., 2011). We found no difference between the proportion of each type of statement and the self-construal priming conditions. Statements were on average 60% individual statements, 32% collective statements and 9% relational statements, regardless of self-construal prime. Although our priming and manipulation check task has commonly been used for cross-cultural research in the past, a lack of a measureable effect could be due to many reasons: 1) the self-construal priming

manipulation failed[11], 2) the "Who am I" manipulation check failed, and/or 3) the manipulation check was too long after the priming task (after the games and DV's) that there was no longer a measureable effect. Nonetheless, we continue with the main analysis keeping these possibilities in mind.

Dependent variables

To test our main hypothesis, the second-order difference score (i.e., [Game 1 – Game 2] – [Game 3 – Game 4]) regarding the three DV's (moral outrage, anger, punishment) was analyzed. Therefore, scores represented the relative care for the injustice of in-group members compared to just situation and compared to injustice toward the self. However, for the punishment variable, 68% of participants did not punish at all and only 11% had a positive difference score; therefore, this variable was not suitable to analyze as a dependent variable. Anger and moral outrage as raw scores correlated *rs* > .84 and the second order difference scores correlate *r* = .62. Given the high correlation between the scales, we averaged the two into one variable, which is Emotional Outrage, to test our hypothesis.

Primary analysis

In our preregistration document, we specified a 2x3 ANOVA. However, on closer examination, a significant amount of the variance in the DV anger, 33%, was due to differences between groups, while 67% amount of variance is due to differences between participants within groups. Therefore, we proceeded to test our hypothesis using multilevel modeling which can take into account the nested nature of the data (i.e., treat group as a random variable on level-2).

We found no main effects of the group membership manipulation (*B*=0.05, *SE*(*B*)=0.23, *p*=.82) nor of either of the self-construal priming conditions (individual prime: *B*=-0.20, *SE*(*B*)=0.22, *p*=.36, relational prime: *B*=-0.18, *SE*(*B*)=0.23, *p*=.43) on emotional outrage.

[11] Some participants did not follow all of the directions in the priming task, by not circling all the words however there was no difference between these participants and the others in the "Who I am?" statements or other variables.

Furthermore, no two-way interaction between them (group membership x individual prime: *B*=0.12, *SE*(*B*)=0.32, *p*=.72, group membership x relational prime: *B*=0.04, *SE*(*B*)=0.32, *p*=.90) was found.

However, when we enter the previously designated potential moderators VS and covariate OS as standardized variables, we found the following: a main effect of OS (*B*=0.28, *SE*(*B*)=0.09, *p*=.003), and a marginal three way interaction between group membership, relational self-construal prime, and victim sensitivity (*B*=0.82, *SE*(*B*)=0.53, *p*=.065) on emotional outrage (see Table 2 for the results of the entire model). The plot of this interaction (see Figure 1) shows that looking specifically at the three-way interaction with the relational prime vs no prime contrast, those high in victim sensitivity show less relative emotional outrage when there was a match between the victim being in the relational group and a relational self-construal prime. However, these results were not significant according to conventional levels, so interpretation of the results should be treated cautiously.

Table III-2

Results for the analysis of moderators VS and covariate OS

	B	*SE(B)*	df	*t*	Sig.
Intercept	-0,30	0,53	141,00	-0,56	0,575
Group type	0,50	0,81	141,00	0,61	0,543
Relational prime contrast	1,17	1,01	141,00	1,16	0,248
Individual prime contrast	-0,25	0,74	141,00	-0,34	0,737
Victim Sensitivity	-0,14	0,19	141,00	-0,75	0,457
Observer Sensitivity	0,28	0,09	141,00	3,08	0,003
Group type by Relational prime contrast	-2,37	1,31	141,00	-1,80	0,074
Group type by Individual prime contrast	-0,42	1,13	141,00	-0,38	0,709
Group type by Victim Sensitivity	-0,15	0,28	141,00	-0,54	0,588
Relational prime contrast by Victim Sensitivity	-0,44	0,34	141,00	-1,31	0,192
Individual prime contrast by Victim Sensitivity	0,03	0,26	141,00	0,13	0,897
Group type by Relational prime contrast by Victim Sensitivity	0,82	0,44	141,00	1,86	0,065
Group type by Individual prime contrast by Victim Sensitivity	0,18	0,38	141,00	0,48	0,633

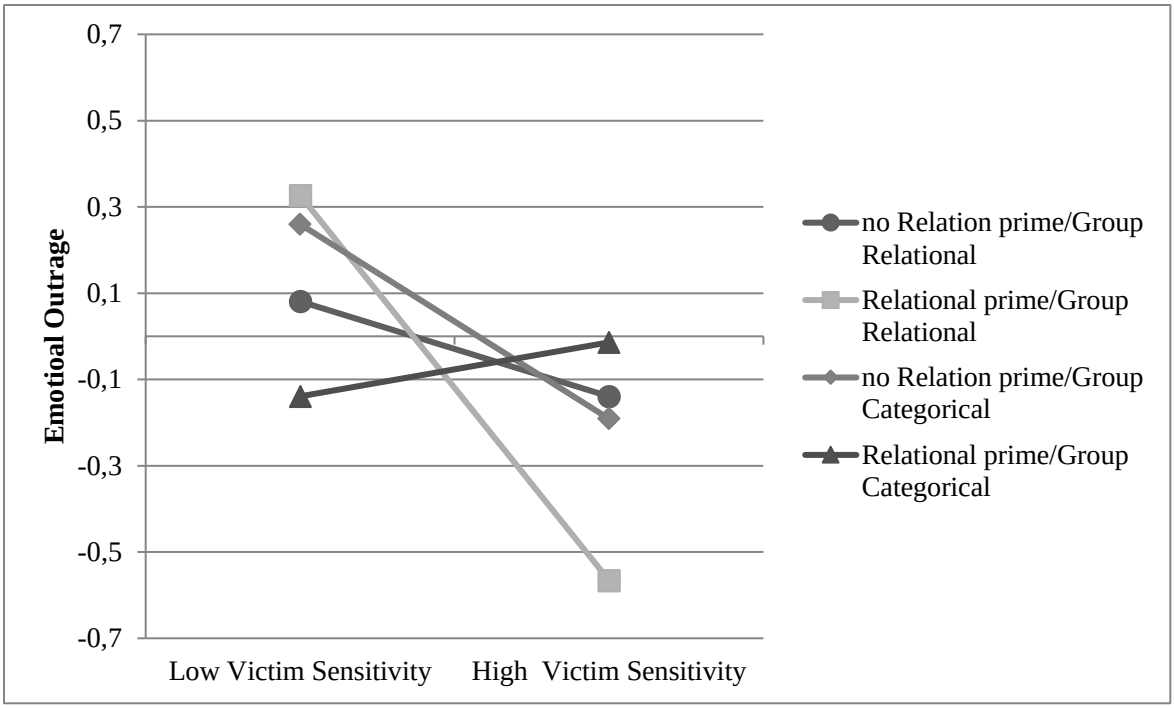

Figure III-1 Three-way interaction between group membership, relational self-construal prime, and victim sensitivity.

Exploratory analysis

Due to the fact that we do not have evidence that our manipulation of self-construal truly worked, we did an exploratory analysis to look at the measured self-construal (independent self-construal and interdependent self-construal) rather than the manipulated self-construal as independent variables. Looking again at the nest model as above, including OS and VS as covariate, we found a marginally significant interaction between group member type and interdependent self-construal scale (B=-0.24, $SE(B)$=0.13, p=.063), but not with group member type and independent self-construal scale (B=0.03, $SE(B)$=0.13, p=(.84), (see Table 3 for the results of the entire model). The plot of this interaction (see Figure 2) shows that for those in the categorical victim condition, when interdependent self-construal was high, there was more relative emotional outrage compared to those low on interdependent self-construal. This simple effect was marginally significant (B=-0.16, $SE(B)$=0.09, p=.094). For participants in the relational victim condition there was no significant difference between those high or low in interdependent self-construal (B=-0.05, $SE(B)$=0.09, p=.58).

Table III-3

Results for the analysis of group type and independent and interdependent self-construal with covariates OS and VS

	B	*SE(B)*	df	*t*	Sig.
Intercept	-0,12	0,09	141,00	-1,31	0,191
Group type	0,08	0,13	141,00	0,65	0,518
Victim Sensitivity	0,26	0,07	141,00	3,64	0,000
Observer Sensitivity	-0,11	0,07	141,00	-1,58	0,117
Independent self-construal	-0,09	0,09	141,00	-0,95	0,344
Interdependent self-construal	0,06	0,09	141,00	0,72	0,475
Group type by Independent self-construal	0,03	0,13	141,00	0,21	0,837
Group type by Interdependent self-construal	-0,24	0,13	141,00	-1,87	0,063

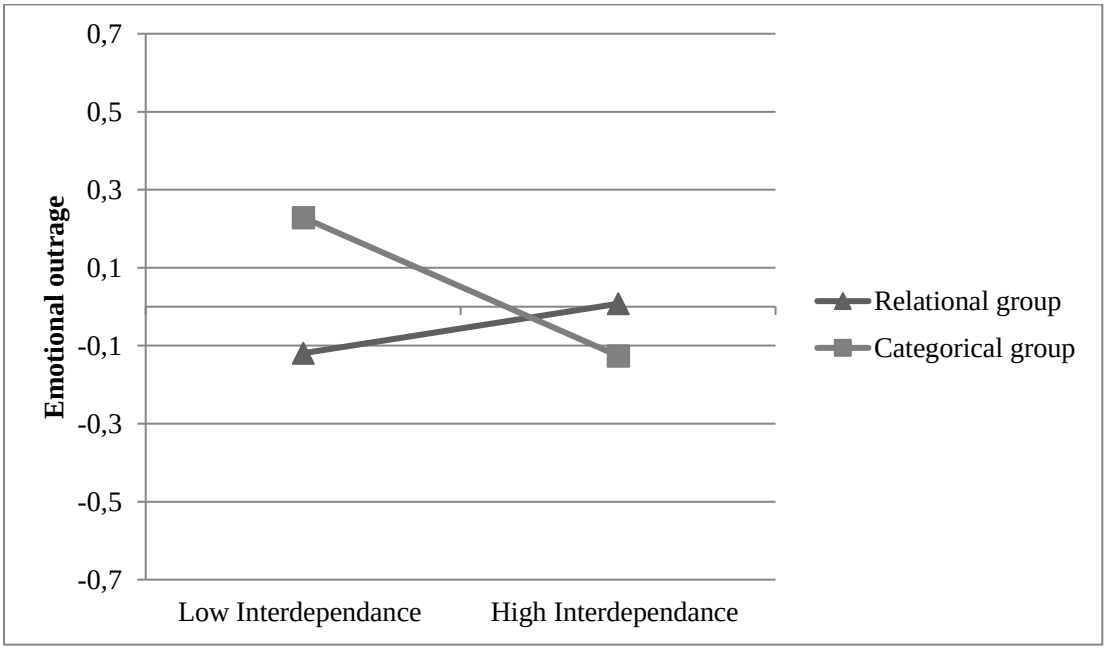

Figure III-2 Interaction between group member type and interdependent self-construal scale.

Discussion

We conducted a between-subject experimental study in China that manipulated group membership: categorical group member vs. relational group member, and self-construal: individual self vs. relational self vs. neutral. First, we measured personality variables, which was followed by a minimal group paradigm and priming procedure. A trust game then served to confront participants with an unfair event, to which we measured their emotional reaction.

It was important to manipulated self-construal, not only because it leads to less ambiguity about the casual relationship, but also because this method was clearer in the distinction between the relational self and the categorical self. However, we did not find the expected interaction between group membership and primed self-construal. This could be due to the priming procedure itself. There is evidence for the prime being less then successful in the results of the manipulation check, that is, the "Who I am?" statement task did not show the expected variation by priming condition. This could be due to the fact that these procedures have not been tested fully in China but also due to the limitations of priming culture in general. It is known from the literature that priming effects experimentally induced are qualitatively different from the effect culture has on psychological behavior in several ways; importantly, priming effects are fleeting, whereas the cultural influences are stable and lasting (Suh, Diener, & Updegraff, 2008). With this limitation in mind it was important to follow up by testing our main hypothesis with measured self-construal.

With this analysis we found some support for our main hypothesis: participants low in interdependent self-construal. In other words, when the relational self is relatively unimportant to an individual, there was more emotional outrage when the victim was from their categorical group compared to when the victim was from their relational group. This supports the theory that we predict more moral outrage when there is a match between one's self-construal, that is, the relational self, and group type, that is the relational group member harmed. However, this also points to a complicated relationship between the two scales for self-construal and the further differentiation between the relational and categorical self-construal. The measured self-construal scales are not differentiated between individual self, relational self, and categorical self. However, we can presume from our results, in line with the hypothesis presented in the literature (Bewer & Chean 2007), that the low end of the interdependent self scale taps into the categorical self. That is, when the interdependent self-construal is less relevant, the self for whom categorical groups are relevant is more important.

However, much more research is need in the area of measurement of self-construal and the relationship between the different types of self-construal, so as to clarify if they are orthogonal concepts.

In addition to our main hypothesis we also looked at a secondary research question regarding the moderating role of victim sensitivity. We found a marginal effect of victim sensitivity, however, in the opposite direction as predicted. Participants high in victim sensitivity showed less outrage when there was a match between the relational self and relational group type. Therefore, rather than expanding the self to include relevant others, people high in victim sensitivity showed more concern for their own injustice relative to others when the victim's group matched the primed self-construal. Future research should look into the issue of when victim sensitive people include groups in the self. It is possible that the features of minimal groups have the opposite effect as real groups for people high in victim sensitivity.

Although this study raises important question about manipulating self-construal and how that relates to measures of independent and interdependent self-construal, this study contributed to the exploration of new methods in the cross-cultural and self-construal literature. Very few studies have attempted priming procedures in China, and, to our knowledge, previous research has not attempted to apply the minimal group paradigm to relational groups. Future research should expand on this by looking at additional cultural samples, especially because the effects of cultural primes when there are differences in national culture is not yet fully understood.

An additional caveat of these results is the reported effects are only marginally significant; therefore we take caution in interpreting them. Future studies should replicate these findings to verify the results.

In conclusion, this research shows that the degree to which one is outraged over the injustice affecting another person contrasted to outrage at one's own injustice is influenced by

the interaction between self-construal and the shared group membership type to the victim. However, more research is needed to strengthen these findings.

References

"When": A meta-analysis of motivational primacy in self-definition. *Journal of Personality and Social Psychology, 83,* 574–591.

Aron, A., Aron, E.N., Tudor, M. and Nelson, G. (1991) Close relationships as including other in the self. *Journal of Personality and Social Psychology*, *60,* 241-253.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*, 122-142.

Brewer, M. B., & Chen, Y.-R. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological Review*, *114*, 133–151.

Brewer, M. B., & Gardner, W. (1996). Who is this "We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology*, *71*, 83–93.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, *1*, 185-216.

Capozza, D., Voci, A., & Licciardello, O. (2000). Individualism, collectivism and social identity theory. *Social identity processes: Trends in theory and research*, Sage. 62-80.

Cross, S. E., Bacon, P. L., & Morris, M. L. (2000). The relational-interdependent self-construal and relationships. *Journal of personality and social psychology*, *78*, 791-808.

Cross, S. E., Hardin, E. E., & Gercek-Swing, B. (2011). The what, how, why, and where of self-construal. *Personality and Social Psychology Review*, *15,* 142-179.

Gaertner, L., Sedikides, C., & Graetz, K. (1999). In search of self-definition: Motivational

Gaertner, L., Sedikides, C., Luke, M., O'Mara, E. M., Iuzzini, J., Jackson, L. E., ... & Wu, Q. (2012). A motivational hierarchy within: Primacy of the individual self, relational self, or collective self?. *Journal of Experimental Social Psychology*, *48*, 997-1013.

Gaertner, L., Sedikides, C., Vevea, J., & Iuzzini, J. (2002). The "I", the "We", and the

Gardner, W. L., Gabriel, S., & Hochschild, L. (2002). When you and I are" we," you are not

threatening: the role of self-expansion in social comparison. *Journal of personality and social psychology*, *82*, 239.

Gardner, W.L., Shira, G. & Lee, A.Y. (1999). "I" value freedom, but "we" value relationships: Self-construal priming mirrors cultural differences in judgment. *Psychological Science*, *10*, 321-326.

Gollwitzer, M., Braun, J., Funk, F., & Süssenbach, P. (2016). People as intuitive retaliators. *Social Psychological and Personality Science*, 7, 521–529.

Gollwitzer, M., Schmidthals, K., & Pöhlmann, C. (2006). Relationalitäts - Kontextabhängigkeits -Skala (RKS): Entwicklung und erste Ansätze zur Validierung. (Berichte aus der Arbeitsgruppe "Verantwortung, Gerechtigkeit, Moral" Nr. 161). Trier: Universität Trier, Fachbereich I – Psychologie.

Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J., & Baer, A. (2005). Asymmetrical effects of justice sensitivity perspectives on prosocial and antisocial behavior. *Social Justice Research*, *18*, 183-201.

Hannover, B., & Kuhnen, U. (2004). Culture, context, and cognition: The semantic procedural interface model of the self. *European review of social psychology*, *15*, 297-333.

Hong, Y. Y., Morris, M. W., Chiu, C. Y., & Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American psychologist*, 55, 709.

Kashima, E.S., Hardie, E.A., Wakimoto, R., & Kashima, Y. (2011). Culture- and gender-specific implications of relational and collective contexts on spontaneous self-descriptions. Journal of Cross-Cultural Psychology, 42, 740-785.

Kuhn, M.H., & McPartland, T.S. (1954). An empirical investigation of self-attitudes. American Sociological Review, 19(1), 68-76.

Kühnen, U., Hannover, B. & Schubert, B. (2001). The semantic-procedural interface model of the self: The role of self-knowledge for context-dependent versus context-independent modes of thinking. *Journal of Personality and Social Psychology, 80*, 397-409.

Leach, C. W., Van Zomeren, M., Zebel, S., Vliek, M. L. W., Pennekamp, S. F., Doosje, B., et al. (2008). Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology, 95*, 144 – 165.

Magraw-Mickelson, Z., & Gollwitzer, M. (2018). Relational and Group Collective Self Responses to Observed Victimization Across Cultures. *Social Justice Research, 31*, 113-132.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98*, 224-253.

Montalan, B., Lelard, T., Godefroy, O., & Mouras, H. (2012). Behavioral investigation of the influence of social categorization on empathy for pain: a minimal group paradigm study. *Frontiers in psychology*, *3*, 389.

Oyserman, D. (1993). The lens of personhood: Viewing the self and others in a multicultural society. *Journal of Personality and Social Psychology*, *65*, 993-1009.

Oyserman, D., & Lee, S. W. S. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological Bulletin, 134*, 311-342.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 3–72.

Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). Us and them: social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology*, *59*, 475.

Perreault, S., & Bourhis, R.Y. (1998). Social identification, interdependence and discrimination. *Group Processes & Intergroup Relations, 1,* 49-66.

Platow, M.J., McClintock, C.G., & Liebrand, W.B.G. (1990). Predicting intergroup fairness and ingroup bias in the minimal group paradigm. *European Journal of Social Psychology, 20*, 221-239.

primacy of the individual self, motivational primacy of the collective self, or contextual primacy? *Journal of Personality and Social Psychology,* 76, 5–18.

Reynolds, K. J., Turner, J. C., Alexander Haslam, S., Ryan, M. K., Bizumic, B., & Subasic, E. (2007). Does personality explain in-group identification and discrimination? Evidence from the minimal group paradigm. *British Journal of Social Psychology*, *46*, 517-539.

Rohrmann, S., Hodapp, V., Schnell, K., Tibubos, A. N., Schwenkmezger, P., & Spielberger, C. D. (2013). Das State-Trait-Ärgerausdrucks-Inventar-2 (STAXI-2). Deutschsprachige Adaptation des State-Trait Anger Expression Inventory-2 (STAXI-2) von Charles D. Spielberger. Bern: Verlag Hans Huber.

Rubini, M., Moscatelli, S., & Palmonari, A. (2007). Increasing group entitativity: Linguistic intergroup discrimination in the minimal group paradigm. *Group Processes & Intergroup Relations*, *10*, 280-296.

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity. *European Journal of Psychological Assessment*, *21*, 202–211.

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin*, *20,* 580-591.

Singelis, T. M., Bond, M. H., Sharkey, W. F., & Lai, C. S. Y. (1999). Unpackaging culture's influence on self-esteem and embarrassability: The role of self-construals. *Journal of Cross-Cultural Psychology*, *30*, 315-341.

Suh, E. M., Diener, E. D., & Updegraff, J. A. (2008). From culture to priming conditions: Self-construal influences on life satisfaction judgments. *Journal of Cross-Cultural*

Psychology, 39, 3-15.

Tajfel, H., & Turner. J. C. (1979). An integrative theory of intergroup conflict. In W. Austin & S. Worchel (Eds.), Social psychology of intergroup relations (pp. 33–47). Chicago: Nelson.

Taras, V., Sarala, R., Muchinsky, P., Kemmelmeier, M., Singelis, T. M., Avsec, A., ... & Hardin, E. E. (2014). Opposite ends of the same stick? Multi-method test of the dimensionality of individualism and collectivism. *Journal of Cross-Cultural Psychology, 45*, 213-245.

Trafimow, D., Triandis, H. C., & Goto, S. G. (1991). Some tests of the distinction between the private self and the collective self. *Journal of Personality and Social Psychology, 60*, 649-655.

Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.

Van den Boos, K., van Veldhuizen, T.S., & Au, A.K.C. (2014). Counter cross-cultural priming and relative deprivation: The role of individualism – collectivism. *Social Justice Research, 28*, 52-75.

Van den Bos, K., Brockner, J., Van den Oudenalder, M., Kamble, S.V., & Nasabi, A. (2013). Delineating a method to study cross-cultural differences with experimental control: The voice effect and countercultural contexts regarding power distance. *Journal of Experimental Social Psychology, 49*, 624-634.

Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism. *Advances in group processes, 16*, 161-197

Yamagishi, T., Jin, N., & Miller, A. S. (1998). In-group bias and culture of collectivism. *Asian Journal of Social Psychology, 1*, 315-328.

Yuki, M. (2003). Intergroup Comparison versus Intragroup Relationships: A Cross-Cultural Examination of Social Identity Theory in North American and East Asian Cultural Contexts. *Social Psychology Quarterly, 66*, 166-183.

Appendix A

Text for priming conditions

Before you will get to see the results of several games, you will participate in the following short task.

Please read the text below carefully.

"A day in the city"

"I go to the city often. My anticipation fills me as I see the skyscrapers come into view. I allow myself to explore every comer, never letting an attraction escape me. My voice fills the air and street. I see all the sights, I window shop, and everywhere I go I see my reflection looking back at me in the glass of a hundred window. At nightfall I linger, my time in the city almost over. When finally I must leave, I do so knowing that I will soon return. The city belongs to me."

(Prime of the individual self)

"My partner and I go to the city often. Our anticipation fills us as we see the skyscrapers come into view. We allow ourselves to explore every comer, never letting an attraction escape us. Our voice fills the air and street. We see all the sights, we window shop, and everywhere we go we see our reflection looking back at us in the glass of a hundred window. At nightfall we linger, our time in the city almost over. When finally we must leave, we do so knowing that we will soon return. The city belongs to us."

(Prime of the relational self)

"Vehicles go to the city often. Cars fill the streets when skyscrapers come into view. Taxis are allowed in every comer in the city, taxis are on all the city streets. Trucks deliver goods to shops, many cars zoom by and the buses are full. Vehicles horns fill the air and street, the taxis are often the loudest. At nightfall there are fewer cars, but always some taxis. When there are finally less cars and trucks on the street, it is only temporary the cars, trucks and taxis will soon return. The city belongs to the vehicles."

(Prime of the relational self)

Please read the text above once again and carefully circle *all* the pronouns you can find in the text.

(Prime of the individual self) (Prime of the relational self)

Please read the text above once again and carefully circle *all* the vehicle words you can find in the text.

(No prime)

Part 2: Methods

IV. Chapter 4: Survey Mode and Data Quality: A Cross-Cultural Comparison of Careless Responding Across Three Modes

Zoe Magraw-Mickelson, Huan Wang and Mario Gollwitzer

Abstract

Much psychological research depends on participants' diligence in filling out materials such as tests or surveys. However, not all participants are motivated to respond attentively, which leads to unintended issues with the quality of the data. Our question is: how do different modes of data collection - paper/pencil, computer/web-based, and smartphone - affect participants' diligence vs. "careless responding" tendencies and, thus, the data quality? Results from prior studies suggest that different modes of data collection produce a comparable prevalence of careless responding tendencies. However, as technology develops and data are collected with increasingly diverse populations, this question needs to be readdressed and taken further by looking at cultural differences. The present research examined the effect of survey mode on careless responding across three waves in a repeated-measures design. Following recommendations in the literature, we computed a careless responding index as a composite of eight indicators that capture aspects of a participant's inattentiveness. In a sample of working adults from China, we found that participants were significantly more careless when completing computer/web-based survey materials than in paper/pencil mode. In a sample of German students, participants were significantly more careless when completing the paper/pencil mode compared to the smartphone mode. In a sample of Chinese-speaking students, we found no difference between the modes. This paper will discuss why these results deviate from past findings that investigated study modes and hypothesize about potential cross-cultural differences.

Introduction

Psychological research often depends on self-report surveys to collect data, which depends on participants giving meaningful answers. Participants giving answers without regarding the content of the questions is a source of measurement error that can lead to unusable data or even inaccurate conclusions. Furthermore, psychological research is expanding in two significant ways: technological speaking, there has been a distinct shift from paper and pencil surveys to those done on computer/web platforms, to the increasing use of mobile phone platforms. With this shift, some studies suggest that these are equally effective platforms (Casler, Bickel, & Hackett, 2013; De Beuckelaer & Leievens, 2009; Dodou & de Winter, 2014). However, as humans' familiarity with and the novelty of technology is ever shifting, the quality of data generated from different platforms warrants reexamination overtime. Second, while most studies that appeared in mainstream psychological journals were based on a homogeneous samples from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (see Henrich, Heine, and Norenzayan, 2010), an increasing number of studies now include subsamples from different countries or cultures, or try to replicate their findings in these countries/cultures in order to scrutinize their external validity (e.g., Aknin et al., 2013).

Developments in technology had facilitated many researches on self-report data quality such as rudimentary data screening methods, i.e. straightlining or string responses (Zhang & Conrad, 2014) and cross platforms comparison (Casler, Bickel, & Hackett, 2013; Yun & Trumbo, 2000). However, there is a research gap on the cultural influences of "careless responding". In a cross-cultural context of Germany and China, this article answer the question: how do different modes of data collection - paper/pencil, computer/web-based,

and smartphone – interact with cultural values to affect participants' diligence vs. "careless responding" tendencies and, thus, the data quality? .

Careless Responding: Definitions and Identification Strategies

Careless responding (CR) is defined as data that "doesn't reflect respondents' true levels of the construct purportedly being measured" (Meade & Craig, 2012). In the literature, this has also been referred to as *insufficient effort responding* (Bowling et al., 2016; Curran, 2016, 2016; Huang Curran, Keeney, Poposki, & DeShon, 2012), *inattentive responding* (Maniaci & Rogge, 2014), and *satisficing* (Fang, Wen, & Prybutok, 2014), all of which describe a response style that is random in regard to the item's content. This is conceptually different than *response bias*, which is a nonrandom source of error. Although random (i.e., unsystematic) measurement error primarily affects the reliability of a measure, it can, under certain circumstances (e.g., large deviations from the mean), also increase the likelihood of systematically biased results (e.g., regression to the mean; cf. Huang, Liu, & Bowling, 2014; see also Gollwitzer, Christ, & Lemmer, 2014).

Typical methods of data screening, such as looking at response time and eliminating cases with impossibly fast times can catch some of the most extreme cases of careless responding; however, these methods can be fallible because they tend to overlook cases in which responses were careless, albeit within plausible response time ranges. Our approach is to look at an array of careless responding (CR) indices instead of a single indicator. This is advantageous for many reasons. First, any single method would miss other types of CR. This is highlighted in the fact that there is more than one style of responding that does not reflect the questions at hand. Second, many methods are designed to catch responders who are equally careless throughout the whole survey, in response to each and every question, which is often not the case. For example, a common method is to use single item questions, such as "use me" prompts (e.g., "In your honest opinion, should we use your data in our analyses in this study?") or bogus questions (e.g., "I sleep less than one hour per night."). Needless to say,

such approaches are likely to produce a type-1 error (i.e., falsely flagging a response as "careless") or a type-2 error (i.e., failing to flag CR as such).

A review of the literature came up with twelve common measures that are used to calculate a careless responding index: response time, long string max, long string mean, Mahalanobis distance, person total correlation, even-odd consistency, psychometric synonyms, psychometric antonyms, self-reported effort, a self-reported single 'use me' item, bogus questions and instructed response items (Bowling et al., 2016; Curran, 2016; DeSimone, Harms, & DeSimone, 2014; Huang et al., 2012; Huang, et al., 2014; Huang, Bowling, Liu, & Li, 2015; Johnson, 2005; Maniaci & Rogge, 2014; Meade & Craig, 2012; Ward et al., 2017). We will now take a closer look at these measures and what types of careless responding tendencies they capture.

Response time is one of the most common data screening methods. It is normally calculated by determining the minimum time required to complete a survey or a block of questions. This cutoff value is then used to exclude participants who complete the survey in an impossibly or implausibly short amount of time. This identifies careless responders whose motive is to go through survey materials as quickly as possible.

Long-string analysis: Long-string max and mean is the analysis of string responses, measuring the number of times a respondent answered with an unbroken sequence on a Likert scale. As an index, this can be calculated as the most identical responses in a row across measures, across pages of a survey, or as the mean of the longest responses across pages. This method identifies careless responders who put little or no effort into varying their responses. Slight deviations from a single response value may not be identified by this measure.

Mahalanobis distances is used to identify outliers that arise due to careless responding (Bowling et al., 2016; Meade & Craig, 2012; Ward et al., 2017). By computing the participants distance from the average participants' response pattern, this method

identifies unusual data points relative to other data points in the sample using a multivariate technique.

Individual consistency: this set of indices can be grouped under the assumption that attentive participants will respond in an internally consistent way. There are different methods to determine individual consistency. First, **Psychometric synonyms,** items that have a high inter-person-correlation in the overall data set that are then correlated at the individual level, the logic being that participants who have low or negative correlations were not attentive to the content of alike questions. Similarly, **psychometric antonyms** find the intra-person correlation for the items that are found to be most negatively correlated across the data set. This could, for example, identify participants who ignored reverse coded items. **Person-total-correlation** is another measure of individual consistency that measures each person's consistency of responding with that of all other participants. A negative correlation for this measure identifies individuals who responded in a contrasting way to the pattern set out by all other participants. **Even-odd consistency** splits unidimensional scales into two subscales, even numbered and odd numbered items, and finds the correlation between the average of these two halves, measuring the consistency in answering items of the same scale.

Self-report items: Self-reported effort asks participants to disclose how much effort they put into the task with questions such as "I worked to the best of my abilities in this study." (Meade & Craig, 2012). A **self-reported 'use me' item** is usually a single item similar to self-reported effort in which the participant tells you if they think they were sufficiently attentive. These measures depend on the honesty of the participant and their attentiveness when the question is asked. However, these measures capture the fact that the participant is the one who knows best if they were careless in their responding.

Instructed response items and **bogus questions** are similar in that they have a right or wrong answer and depend on participants reading the text of the question to get the correct answer. Instructed response questions tell the participant which response to select (e.g.,

"Please select answer choice 5 for this question") while bogus questions present a statement with only one plausible answer.

Survey Mode, Cultural Norms, and Accountability

The literature is mixed on how to classify careless responding; some argue that CR is an aspect of personality, with a considerable degree of consistency across time and situations, and meaningful associations with other personality dispositions, such as agreeableness (Bowling et al., 2016, Ward et al., 2017). However, there are also significant environmental factors that influence the amount of attention in the moment, such as survey mode (i.e., paper/pencil, computer-based) and cultural norms.

One plausible assumption is that CR is more likely to occur in situations in which participants feel that they cannot be held accountable for their responses. Thus, CR should be less likely to occur in face-to-face interviews (where accountability is highest) compared to paper/pencil questionnaires that are distributed and collected by a researcher or other authority (where accountability is comparably weaker), and most likely to occur in online surveys that are completed in an entirely private environment (i.e., at home; Ward et al., 2017).

While this hypothesis makes a lot of intuitive sense, the research findings are mixed. For instance, Dodou and de Winter (2014) looked at the effect of survey mode (i.e., paper/pencil and computer-based) on social desirability indicators – which should also be related to accountability concerns (cf. Paulhus, 1984) – in a meta-analysis with 51 primary studies comprising, in total, more than 16,000 participants. Interestingly, these authors found no relation between survey mode and social desirability, which seems to suggest that survey mode is also unrelated to accountability. However, it should be noted that the vast majority of primary studies included in Dodou and de Winter's (2014) meta-analysis was conducted with Western participants, and there are reasons to believe that culture plays a significant role in

social desirability concerns and the degree of accountability that is connected to a particular survey mode. More precisely, cross-cultural research suggests that "saving face" is an important concern in Asian cultures (Oetzel et al., 2001). Thus, one could argue that Asian participants might feel more accountable for their responses when they complete a paper/pencil-based questionnaire compared to a more anonymous online-based survey. Consequently, a higher degree of CR in computer/web surveys compared to paper/pencil questionnaires may be observable in Asian, but not Western, cultures. This hypothesis will be explored in the present study.

New Developments in Collecting Online Survey Data

Most of the priori methodological research on survey mode and data quality has focused on the difference between paper/pencil questionnaires versus computer/web surveys, where "computer" usually meant a stationary machine located in a specific, familiar environment (e.g., one's study). However, with the increasing popularity of smartphones, many people do not even own a stationary machine anymore; more and more people do their online businesses exclusively on their smartphone. Consequently, more and more online surveys are now completed and even specifically designed for being completed on smartphones (Peterson, Griffin, LaFrance, & Li, 2017). This raises the question of whether careless responding is more or equally likely to occur when the online-survey is completed on a smartphone compared to a stationary computer. One plausible assumption could be that situational circumstances are much less standardized and much "noisier" (also in a literal sense) when the survey is completed on a smartphone compared to a stationary computer. Hence, careless responding should be more likely to occur when data are collected via smartphone than via computer. This hypothesis will also be explored in the present study.

The Present Research

The present study will look at careless responding in three samples – a typical Western student sample (German university students), a sample of Chinese speaking students (from

two Chinese universities), and a sample of Chinese working adults – to explore the effects of age and culture on CR. Furthermore, in a within-participant design, we will look at careless responding across three modes: paper/pencil, computer/web-based, and smartphone-based. The within-participant design we adopted for the present research is particularly beneficial because we can directly compare how the participants react to different modes beyond compounding individual differences, as past research has shown a connection between CR and personality variables (Bowling et al., 2016; Ward et al., 2017).

The specific items and scales used in the survey were selected on the basis of the following rationale: first, it was important to have a typical psychological survey that a researcher would use. Second, we included a dimension related to survey engagement. Third, we included personality factors hypothesized to be related to survey engagement. Finally, we wanted to keep the overall length of the survey short so as to not create undue burden on our within-participant design.

A careless responding (CR) index was calculated for each participant and each version on the basis of participants' responses to these scales and items. CR values were then compared between survey modes (i.e., paper/pencil, computer/web, smartphone) and samples (German students, Chinese-speaking students, Chinese working adults). The following hypotheses were tested:

H1: In all three samples, the degree of careless responding will be higher in the smartphone mode compared to the paper/pencil and computer/web mode.

H2: In the Chinese students and the Chinese working adults samples, the degree of careless responding in the computer/web mode is higher than in the paper/pencil mode.

Methods

Participants

This study included three samples from four different locations, in which participants were asked to complete a survey in a directed mode: Paper/pencil, online with a computer,

and with a smartphone[12], about two weeks apart. Each sample was divided into three groups, and the order of modes in which the surveys were completed were systematically varied between groups (so that each mode occurred once at each order position; i.e., a Latin-square design to minimize artificial order effects).

The sample of Chinese working adults (N = 78) were employees at a municipal clerk's office in China. The procedure was carried out by the coordination office, and they received no reward for their participation. The sample of German students (N = 117) were primarily first-year psychology students. The procedure was administered through a mandatory course, and the paper/pencil version of the materials was followed by other unrelated survey measures for which the students received course credits for completing all parts. The Chinese speaking student sample came from two locations: university 1 (N = 74), wherein the procedure was administered through university language courses and second, from university 2 (N = 39), where the surveys were administered by a teacher in their organizational psychology course. Both Chinese-speaking student locations received no reward for participating. In all conditions, participants were told the survey was for university research and that their participation was completely voluntary. Participants were explicitly told that personal information that included email, ID numbers and phone numbers, were only for matching between the modes. As soon as the three modes were matched to one participant, an anonymous code would be used and their personal information was deleted. Furthermore, we assured participants that their responses would not be seen by their employers or course instructors.

Excluded from the analysis were entries which did not follow directions regarding completion mode (completing the smartphone version on the computer/web or vice versa), and therefore entered more than one response for a particular version (Chinese working

[12] One participant in each student sample expressed that they did not own a smart phone and were directed to complete the remaining two modes only.

adults: *N* = 4, Chinese-speaking students: *N* = 22, Germany: *N* = 32). For these participants, only the dataset that was completed first was retained. Included in the calculation and examination of CR but excluded from the analysis of differences between modes were participants who completed only one version (Germany: *N* = 51, Chinese-speaking students: *N* = 42). Demographics of the samples are presented in Table 1.

Table IV-1
Sample information for the four sampling locations.

Age	N	Minimum	Maximum	Mean	Std. Deviation	%female
Chinese working	78	21	56	36,03	9,78	57,7
German students	117	18	49	21,07	4,64	88,8
Chinese students, University 1	74	19	22	20,23	0,79	90,5
Chinese students, University 2	39	20	24	21,13	1,08	25,6

Materials

The base of the survey included demographics (age, gender, education level, tenure/semester in university) as well as the following measures, all complete on six-point Likert scales ranging from 0 ("do not agree at all") to 5 ("agree completely"): *Agreeableness* from Mini- International Personality Item Pool (IPIP) (Donnellan et al., 2006), consisting of four items (e.g., "I sympathize with others' feelings"); *social desirability* with 6 items (e.g., "I sometimes tell lies") from the Balanced inventory of desirable responding (BIDR) Version 6 short form (Stöber et al., 2002, BIDR used in China: Fang et al., 2014); a short *harmony motivation* scale (adapted from Leung, 2010,) which includes *harmony enhancement*, 5 items (e.g., "Being patient and willing to compromise is a show of respect to the other person.") and *disintegration avoidance*, 5 items (e.g., "You should not create conflict. When you have a conflict, you should try to smooth it over and make the other person happy."); the four subscales from the Justice Sensitivity Inventory (Schmitt et al., 2010, i.e., *Victim Sensitivity*, *Observer Sensitivity*, *Beneficiary Sensitivity*, and *Perpetrator Sensitivity*, 10 items each);

justice heuristics (Ambrose & Schminke, 2009), 6 items (e.g., "In general, I can count on my supervisor/instructors to be fair."); *attitudes towards surveys* (Stocké, 2006, adapted), 3 items (e.g., "Participation in surveys is generally useful."); *technology ease of use* (adapted from Saadé & Bahli, 2005,), 4 items (e.g., "Using [this questionnaire / this computer / this smartphone app] for this survey is easy for me to do."); *feelings of anonymity*, 3 items (e.g., "I felt anonymous completing this survey"); and one item for *work stress* ("Overall, the stress I feel at work is acceptable"); *self-reported effort*, 4 items (e.g. "I worked to the best of my abilities in this study.") (Meade & Craig, 2012).

Computing a Careless Responding Index

Because this design included a paper/pencil survey, response time could not be calculated. Even-odd consistency was also not a good fit as an index for this data set because it requires a comparison between the odd and even questions across various sufficiently long univariate scales. This is known to be a limitation of this technique (Curran, 2016). Furthermore, we only included three instructed response items, as recommended by Meade and Craig (2012). These are preferable to bogus questions and because of the short length of our survey, we did not want to annoy participants with too many of this type of questions. Although we collected a Yes/No 'use me' question, we did not include this in the index because of its dichotomous nature.

Therefore, we included eight indicators of careless responding in the composite index. In the following, we will describe the method used to calculate each of the indices.

To calculate **long string max and mean,** we used three pages or blocks of questions that included different measures (Block 1: agreeableness and social desirability; Block 2: attitudes, ease of use, anonymous, effort; Block 3: harmony enhancement). Long string Max was the longest row of identical answers across all three blocks; Long string Mean was the average of long string answers across three blocks.

Mahalanobis distances. To calculate this measure, we averaged the Mahalanobis distances of 9 scales: Justice sensitivity scales, Agreeableness, Social Desirability, Harmony Enhancement, Disintegration Avoidance, Justice Heuristics. As seen in Table 2, for our data sets the correlation with Mahalanobis distances and other indices is rather low compared to those reported in past studies which report similar measures. This may be due to the distribution of the data, which is, skewed versus an even distribution. Past research has shown that Mahalanobis distances' utility as an indicator is sensitive to the data's distribution (Meade & Craig, 2012).

Psychometric synonyms. Curran (2016) recommended identifying all inter-correlated pairs without repeat items above the cut-off of $r = .60$ between items. However, that strategy depends a great deal on the nature of the sample. Therefore, to establish some consistency across our three samples, we took the ten highest positively correlated item pairs, $rs > .61$. For **psychometric antonyms,** there are also recommendations for negative correlation cut-offs values $r = -.20$. However, we again decided to use a set number of pairs for each samples. Here, we used five pairs due to the fact that the survey was not very long and did not include a large number of reverse coded items $rs < -.30$. Once the pairs of items were identified for each sample, the intra-person correlation of those item sets was computed for each participant. When the correlation of synonyms or antonyms was not calculable due to the lack of variance, meaning either in the numerator or dominator, the participant had the same value for all items in the selected set, and the indices for that participant was entered as a missing value.

Person-total-correlation scores were calculated across 60 items (i.e., justice sensitivity scales, agreeableness scale, social desirability scale, harmony enhancement, and disintegration avoidance scales) exactly as described by Curran (2016).

For **self-reported effort**, we used a four-item measure as described above. For a full list of questions, see Appendix A.

For **instructed response items,** we included three questions per wave. An example of the questions is: "Please check 'very accurate' to show you are paying attention". Answers that matched "very accurate" were coded as 0, and all other responses were coded as 1. The values were summed across three questions; therefore the possible range of this measure was 0-3.

For the **self-reported 'use me' item,** we used the single item from (Meade & Craig, 2012): item "In your honest opinion, should we use your data in our analyses in this study?" Yes was coded as 0, and No coded as 1.

To compute the CR index for each person, we multiplied person-total-correlation, psychometric synonyms, and self-reported effort by -1, so that all measures of CR are scaled in the same direction (i.e., higher values indicating a stronger carelessness). Next we *z*-transformed (i.e., standardized) the eight indices separate by culture and then computed an aggregate score by averaging across indices. See Table 2 for the correlations between the indices.

Table IV-2a

Correlations Among Variables for Chinese Working Adults Sample

Variable	N	1	2	3	4	5	6	7	8	11
1. Long-string Max	78	1								
2. Long-string Mean	78	,900**	1							
3. Mahalanobis distance	78	-0,088	-0,044	1						
4. Psychometric synonyms	78	-0,118	-0,023	,235*	1					
5. Psychometric antonyms	76	0,134	0,145	-0,063	0,207	1				
6. Person total correlation	78	0,19	,230*	,225*	,397**	,520**	1			
7. Self-reported effort	78	0,168	,255*	-,280*	0,104	0,142	0,036	1		
9. Self-reported 'use me'	78	0,059	0,028	0,214	-0,103	0,101	0,096	-,320**	1	
8. Instructed Response	78	0,145	0,005	-0,058	-,258*	0,061	0,119	0,105	0,031	
Careless Responding Index	78	,597**	,631**	,239*	,389**	,532**	,684**	,426**	0,018	,331**

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Table IV-2b

Correlations Among Variables for German students

Variable		1	2	3	4	5	6	7	8	11
1. Long-string Max	117	1								
2. Long-string Mean	117	,822**	1							
3. Mahalanobis distance	117	0,012	0,117	1						
4. Psychometric synonyms	117	-,215*	-0,136	-0,029	1					
5. Psychometric antonyms	115	-0,018	0,008	0,085	-0,052	1				
6. Person total correlation	117	0,043	0,151	,554**	-,221*	,355**	1			
7. Self-reported effort	116	-0,157	-0,151	-0,142	0,117	0,166	-0,103	1		
9. Self-reported ‘use me’	117	-0,067	-0,078	-0,013	-0,021	0,019	0,005	,210*	1	
8. Instructed Response	117	0,06	0,101	0,1	0,053	-0,051	0,045	0,171	0,094	
Careless Responding Index	117	,395**	,505**	,495**	,184*	,445**	,512**	,303**	0,059	,515**

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Table IV-2c

Correlations Among Variables for Chinese speaking students

Variable		1	2	3	4	5	6	7	8	11
1. Long-string Max	113	1								
2. Long-string Mean	113	,890**	1							
3. Mahalanobis distance	113	-,422**	-,308**	1						
4. Psychometric synonyms	108	0,033	0,093	-0,05	1					
5. Psychometric antonyms	113	,247**	,195*	-0,108	-0,029	1				
6. Person total correlation	113	0,061	0,12	0,171	0,111	,225*	1			
7. Self-reported effort	113	,300**	,247**	-,288**	0,02	,370**	,293**	1		
9. Self-reported ‘use me’	113	,196*	,203*	-0,013	0,077	,222*	0,091	,287**	1	
8. Instructed Response	113	,456**	,430**	-0,11	0,052	,233*	,208*	,215*	0,129	
Careless Responding Index	113	,665**	,686**	-0,049	,347**	,565**	,550**	,558**	,312**	,641**

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Results

Main findings

Ideally, we would have a 3 by 3 design with the three modes (paper/pencil vs. computer vs. smartphone; *within-participants*) by three cultures groups (German students vs. Chinese-speaking students vs. Chinese working adults; *between-participants*); however, due to the manner in which the CR index was calculated, specifically, the centering within each culture, the absolute value of the index is not comparable across cultures. For this reason, we instead ran one-factorial mixed models for each country, individually. Descriptive statistics are reported in Table 3.

We looked at the CR index to investigate the contrast of our main hypothesis we included two Helmert-coded contrasts: (1) paper/pencil = -1, smartphone = 2, and computer/web = -1; (2): paper/pencil =-1, smartphone = 0, and computer/web = 1. Thus, the first contrast compares the smartphone mode to the two other modes (as specified in Hypothesis 1), and the second contrast compares paper/pencil against computer/web (as specified in Hypothesis 2). Given that both Hypotheses were directional, the contrasts were tested for statistical significance on a 5% level (one-tailed).

Chinese working adults.

H1: We did not find evidence for H1, responses were not more careless in the smartphone mode compared to the other two modes in the Chinese working adult sample, B=-.02, $SE(B)$=.02, p=.44.

H2: We found evidence for H2, responses were more careless in the computer/web mode compared to paper/pencil mode in the Chinese working adult sample, B=.07,

SE(*B*)=.03, *p*=.025 (see Figure 1).

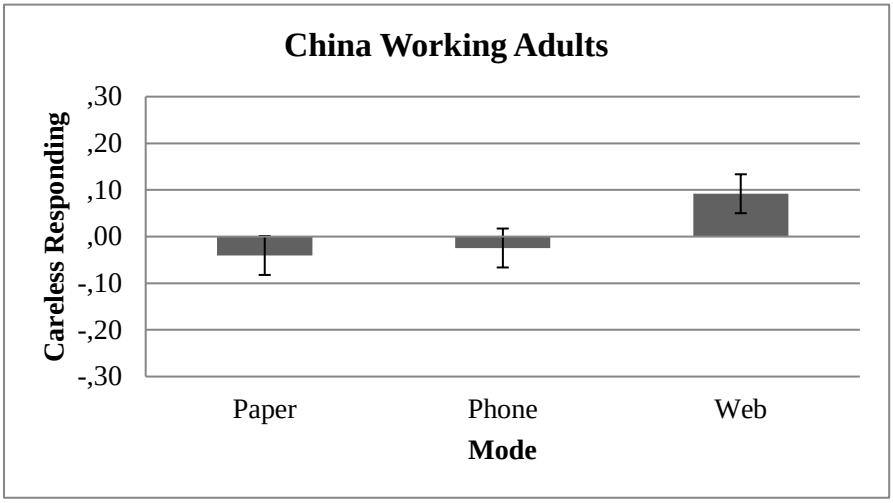

Figure IV-1 Differences between mode in careless responding for the Chinese working adults sample

Chinese students.

H1: In the Chinese student sample, we also did not find evidence for H1, responses were not more careless in the smartphone mode compared to the computer/web mode and the paper/pencil mode, *B*=.00, *SE*(*B*)=.02, *p*=.81.

H2: In the Chinese student sample, we did not find evidence for H2, responses were not more carless in the computer/web mode compared to paper/pencil mode, *B*=.02, *SE*(*B*)=.03, *p*=.54 (see Figure 2).

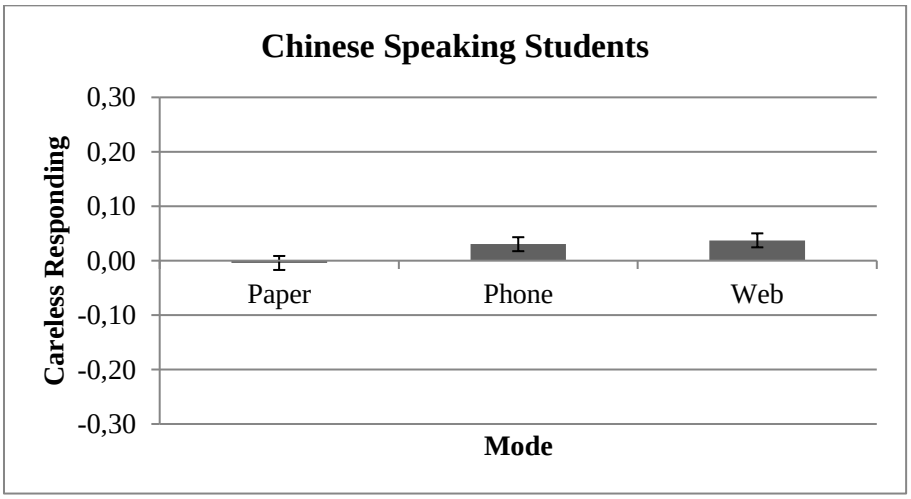

Figure IV-2 Differences between mode in careless responding for the Chinese speaking student sample

German students.

H1: In the German student sample, we also did not find evidence for H1, in fact we found the opposite: responses were even *more* careless when made on the computer/web and the paper/pencil mode compared to the smartphone, B=-.02, $SE(B)$=.01, p=.038. To follow-up on this effect, we tested the difference between smartphone and computer/web as well as the difference between smartphone and paper/pencil, separately. We found a significant difference in paper/pencil compared to smartphone, B=-.10, $SE(B)$=.04, p=.022, but not computer/web compared to smart phone, B=-.06, $SE(B)$=.05, p=.231.

H2: Although not our initial hypothesis we did the same H2 analysis for the German students as the other two samples and found no evidence in difference in the degree of careless responding in the computer/web mode compared to the paper/pencil mode, B=-.03, $SE(B)$=.03, p=.28 (see Figure 3).

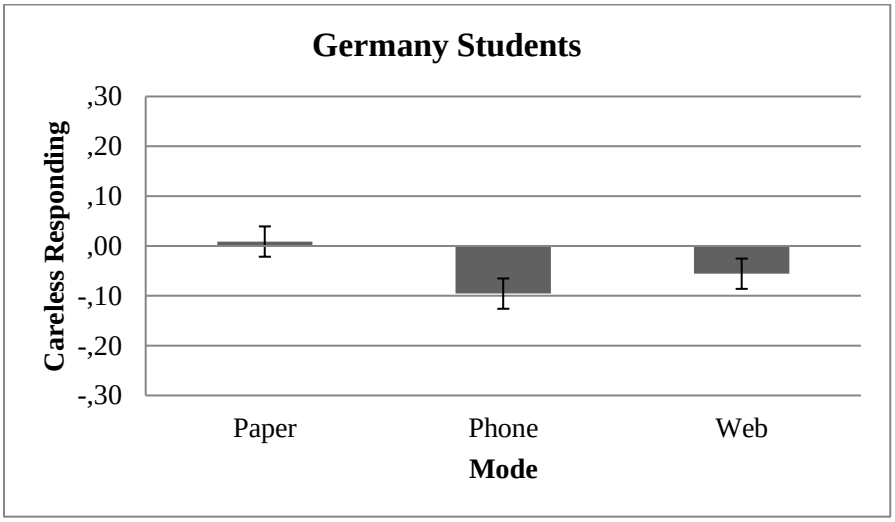

Figure IV-3 Differences between mode in careless responding for the German student sample

Individual indices

Chinese working adults.

Looking at the different indices individually, in line with H2 we found that, for psychometric antonyms and person-total-correlation, participants showed significantly more

carelessness in the computer/web mode compared to the paper/pencil mode (*B*=.14, *SE*(*B*)=.07, *p*=.02, and *B*=.13, *SE*(*B*)=.06, *p*=.02, respectively).

Chinese students.

There were no significant differences between the modes and any of the indices for the Chinese student sample.

German students.

Counter to H1 for the measure of psychometric synonyms, participants showed significantly more carelessness in the paper/pencil and computer/web mode compared to the smartphone mode, *B*=-.09, *SE*(*B*)=.05, *p*=.04. In self-reported effort, contrary to H2, German students were significantly more careless in the paper/pencil mode compared to computer/web, *B*=-.13, *SE*(*B*)=.07, *p*=.03. The instructed response items and 'use me' item significance was not calculated because there were so few in the German sample who missed these questions, and in some modes none at all.

Additional exploratory findings

In an exploratory fashion, we investigated gender differences in careless responding as a function of sample and survey mode. Broken down by sample, in the Chinese working adults sample, men were significantly more careless than women in the paper/pencil mode condition, *B*=-.38, *SE*(*B*)=.09, *p*<.001, in the computer/web mode condition, *B*=-.51, *SE*(*B*)=.12, *p*<.001, and in the smartphone condition, *B*=-.21, *SE*(*B*)=.11, *p*=.03. In both student samples, men were significantly more careless than women in the computer/web mode (German students: *B*=-.51, *SE*(*B*)=.10, *p*=.002; Chinese speaking students: *B*=-.34, *SE*(*B*)=.16, *p*=.017), but not in the other two modes, *p*s > .12.

Table IV-3

Descriptive Statistics indices, in three samples

	Chinese working adults sample N (listwise) = 214				**Chinese speaking students sample** N (listwise) = 172				**German students sample** N (listwise) = 159			
	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD
Longstring Mean	4,94	2,00	10,00	1,63	4,01	2,00	11,33	1,44	2,85	1,67	5,00	0,59
Longstring Max	7,03	2,00	14,00	2,91	5,64	2,00	14,00	2,61	3,63	2,00	7,00	0,96
Mahalanobis D.	7,08	0,87	23,29	4,63	6,85	0,94	32,51	4,93	7,03	2,31	23,96	3,30
Psychometric synonyms	0,69	-0,31	1,00	0,30	0,74	-0,35	1,00	0,25	0,71	-0,22	1,00	0,24
Psychometric antonyms	-0,66	-1,00	0,84	0,36	-0,46	-1,00	0,82	0,42	-0,75	-1,00	0,41	0,28
Person Correlation	0,54	-0,40	0,88	0,22	0,52	-0,05	0,80	0,18	0,67	0,05	0,87	0,15
Self-reported Effort	4,36	1,00	6,00	0,74	4,63	2,75	6,00	0,81	5,13	3,25	6,00	0,57
Self-reported ‘use me’	0,08	0,00	1,00	0,28	0,07	0,00	1,00	0,26	0,02	0,00	1,00	0,16
Instructed Response Items	0,48	0,00	3,00	0,68	0,47	0,00	3,00	0,82	0,03	0,00	1,00	0,17
Careless Responding Index	0,01	-0,92	1,97	0,52	0,01	-0,80	2,44	0,55	-0,05	-0,87	1,64	0,40

Discussion

Past research has shown that careless responding does not reliably differ between different self-report assessment modes. We challenged these results by looking more closely at careless responding tendencies in three different modes (paper/pencil, computer/web, and smartphone) and in three different samples (German students, Chinese students, Chinese working adults). Looking at each of these three samples separately, and using an aggregated careless responding score consisting of eight different indicators, we found that (1) Chinese working adults were more careless when responding on a computer compared to when responding with paper and pencil; (2) German students were more careless responding with paper and pencil compared to smartphone; and that (3) there were no significant differences in careless responding between modes in the Chinese students sample. Thus, only one of our hypotheses (H2) was corroborated by the data, while H1 could not be confirmed. Contrary to the assumptions of past work in careless responding (Ward et al., 2017), we did not find evidence that smartphone surveys were completed with more carelessness than surveys completed on other platforms (H1), in fact it was the opposite with the German student sample. In the following section, we will discuss our findings against the background of other research that has been conducted in this area.

Careless Responding in Chinese Samples

Fang, Wen, and Prybutok (2014) found that university students in China were more likely to engage in satisficing behavior—defined as a suboptimal response style characterized as reduced cognitive effort when responding to a survey—on a computer compared to a paper and pencil survey. These results are consistent with our Chinese working adults sample but not the Chinese student sample. This points to generational and cultural differences in familiarity with different survey modes, in that there are generational differences in the way people interact with survey modes. The student samples used in past studies are represented as the working adults' sample of the present study; this group might have a different relationship

to how students interact with technology today. This is plausible in that age cohorts are growing up with different set of technological norms than just a few years before.

Careless Responding in the German Students Sample

The results of the German sample were unexpected. Previous research has not found a similar pattern; however, we can speculate on the relationship between environment and careless responding. Our hypothesis relied on the assumption that participants would complete the smartphone survey in a more distractible environment. However, this might not have been the case. The German students received the paper and pencil survey in class, even though it was not related to their class work and may have completed it while still amongst their peers. It could be that participants felt more personally accountable when using their smartphone, or were in a more focused environment compared to the paper/pencil and computer/web mode and therefore less careless. Future research should collect more information about the environmental differences of where surveys are completed to better understand the differences between survey modes.

A More Fine-Grained Analysis of Careless Responding

Looking at single indicators of careless responding, we found similar patterns for the difference between the modes for each country; however, these differences only reached significance for person-total-correlation and psychometric antonyms in the Chinese working adults sample and psychometric synonyms and self-reported effort in the German student sample. This further supports the concept of a composite measure to capture different types of CR because looking at only one or a small number of indicators would not have shown the complete picture in the differences between modes in carless responding.

We also found exploratory results related to gender. Although gender itself is a meaningless psychological variable and therefore unable to “explain” our pattern of result, it is important to note we found that men were more careless than women in all modes in the Chinese adult sample and in the computer mode for the other two samples. This could be

related to the fact that women are more agreeable than men (Lehmann, Denissen, Allemand, & Penke, 2013) and low agreeableness has been shown to be related to more careless responding (Bowling et al., 2016, Ward et al., 2017). However it is unclear why this difference would only be in the computer mode for two of the samples. Future research should explore this issue.

Because the different indices are centered on each sample, we did not directly compare the degree of careless responding across samples. However, there is evidence that the two Chinese-speaking samples were more careless. Looking at the descriptive statistics of individual indices, the means (where comparable), the range and standard deviation (see Table 3), the two Chinese samples had worse values compared to the German sample. For example, the maximum of "psychometric antonyms" was higher in the Chinese samples compared to the German sample meaning that, in the Chinese samples, respondents gave more similar answers to reverse coded items.

Limitations

The limitation of this design is that although we can describe differences between modes and cultures, it is hard to look at the psychological mechanisms behind survey differences, using survey questions that depend on the participant's response. We asked questions related to possible sources of differences in careless responding, such as ease of use and anonymity. However, if careless responders are not paying attention to the content of these questions we will not have reliable comparisons. Future research should look into novel ways to combat this issue.

In computing the careless responding index, we took the average of the centered values across eight indices. This is a strength, in that we took into account various types of careless responding, but it is also a limitation in that it assumes all the indices are equally good at identifying careless responders.

Cut-offs for exclusion criteria based on careless responding can be difficult to prescribe and are beyond the scope of this article. Some of the indices have clear dichotomous cut-offs; however, these do not catch all different types of careless responders. Composite indices can be very useful for that; however, more research needs to be done to provide recommendations that researchers can use to identify careless responders. This can be done through simulation studies which can control the percentage of careless responding and vary the patterns of randomness to see how a CR index performs. And with experimental studies that look at careless responding with novel and labor intensive measures which can be compared to a universally applicable index as the one used in this paper.

Conclusion

This research looked at careless responding in a typical research environment. Our results suggest that data quality is not equal across modes of data collection and it is not the same across different cultural and groups. From this rigorous research method, we learn that not all modes are equal, and that participants interact differently with different modes of data collection. More work needs to be done in this area, but the present study is a first, and, arguably, important step towards understanding situational factors, cultural factors, and gender in how careful participants are when completing surveys.

References

Aknin, L. B., Barrington-Leigh, C. P., Dunn, E. W., Helliwell, J. F., Burns, J., Biswas-Diener, R., ... & Norton, M. I. (2013). Prosocial spending and well-being: Cross-cultural evidence for a psychological universal. Journa*l of Personality and Social Psychology, 104*, 635.

Ambrose, M. L., & Schminke, M. (2009). The role of overall justice judgments in organizational justice research: a test of mediation. *Journal of Applied Psychology*, *94*, 491.

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, *111*, 218-229.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*, 2156-2160.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: a construct validation of a measure. *Journal of applied psychology*, *86*, 386.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4-19.

De Beuckelaer, A., & Lievens, F. (2009). Measurement equivalence of paper‐and‐pencil and Internet organisational surveys: A large scale examination in 16 countries. *Applied Psychology, 58*, 336-361.

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, *36*, 171-181.

Dodou, D., & de Winter, J. C. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior*, *36*, 487-495.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment, 18*, 192.

Fang, J., Wen, C., & Prybutok, V. (2014). An assessment of equivalence between paper and social media surveys: The role of social desirability and satisficing. *Computers in Human Behavior, 30*, 335-343.

Gollwitzer, M., Christ, O., & Lemmer, G. (2014). Individual differences make a difference: On the use and the psychometric properties of difference scores in social psychology. *European Journal of Social Psychology*, *44*, 673-682.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, *33*, 61-83.

Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology*, *30*, 299-311.

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*, 99-114.

Huang, J. L., Liu, M., & Bowling, N. A. (2014). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*, 828.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of research in personality*, *39*, 103-129.

Lehmann, R., Denissen, J. J., Allemand, M., & Penke, L. (2013). Age and gender differences in motivational manifestations of the Big Five from age 16 to 60. *Developmental psychology, 49*, 365.

Leung, K., Brew, F. P., Zhang, Z. X., & Zhang, Y. (2010). Harmony and conflict: A cross-cultural investigation in China and Australia. *Journal of Cross-Cultural Psychology*. 22.

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61-83.

Meade, A.W. & Craig, S.B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437- 455.

Oetzel, J., Ting-Toomey, S., Masumoto, T., Yokochi, Y., Pan, X., Takai, J., & Wilcox, R. (2001). Face and facework in conflict: A cross-cultural comparison of China, Germany, Japan, and the United States. *Communication Monographs*, *68*, 235-258.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology*, *46*, 598.

Peterson, G., Griffin, J., LaFrance, J., & Li, J. (2017). Smartphone Participation in Web Surveys. *Total survey error in practice*, 203-233.

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice Sensitivity. *European Journal of Psychological Assessment, 21,* 202–211

Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of personality assessment*, *78*, 370-389.

Stocké, V. (2006). Attitudes toward surveys, attitude accessibility and the effect on respondents' susceptibility to nonresponse. *Quality and Quantity,40,* 259-288.

Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, *76*, 417-430.

Yun, G. W., & Trumbo, C. W. (2000). Comparative response to a survey executed by post, e-mail, & web form. *Journal of computer-mediated communication*, 6.

Zhang, C., & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *In Survey Research Methods*, *8*, 127-135

Appendix A

Self-Reported Effort scale

1. I worked to the best of my abilities in this study.

2. I put forth my best effort in responding to this survey.

3. I would be interested in reading about the results of this study.

4. I’m in a hurry right now. (R)

V. Chapter 5: Comparing Classic and Novel Approaches to Measurement Invariance in an Eight Culture Sample

Zoe Magraw-Mickelson, Alejandro Hermida, Chuma Owuamalam, Maas Misha'ari Weerabangsa, and Mario Gollwitzer

Abstract

Measurement invariance (MI) is vital to any comparison of heterogeneous groups. With multiple-group confirmatory factor analysis (MG-CFA), which is the standard practice for testing MI, there are widely acknowledged limitations, especially with a large number of groups for which strict invariance is difficult or impossible to achieve. New methods, specifically Alignment optimization method, give increased flexibility and new opportunities to make comparisons across a large number of groups. This article compares MG-CFA with Alignment method for MI testing with a demonstrative example of eight countries looking at eight different measures. We find with this medium size number of groups, strict MI is problematic; only partial invariance was achievable with MG-CFA and in half the measures, some of the countries were excluded. However, there is considerable possibility for group level comparisons without these drawbacks using Alignment method. We compared the implications of the relative difference between the output means according to the different methods and with additional country level variables. In this comparison, we find that for most measures, there are less significant latent mean differences when using Alignment method vs. partial invariance MG-CFA. Together, this demonstration suggests that Alignment method is a promising solution to give more confidence to the comparability of measures in multi-group comparison research.

Introduction

The ability to compare groups is a bedrock principle of any cross-cultural research projects or any research that wants to compare heterogeneous groups, be that across nations, genders, age groups, or time points. Researchers depend on statistical methods to answer questions of comparability, specifically measurement invariance, which is to ensure that participants with the same true-score on a latent variable respond in a similar manner to a manifestation of that variable, that is, without any systematic biases due to group memberships, age, or measurement occasion. Although there has been growing recognition of the importance of establishing measure invariance (MI), there are still challenges to be met for execution, namely what is acceptable as "partial invariance," and what to do if it is not established? Multiple-group confirmatory factor analysis (MG-CFA) is the most common method for testing MI, but this method has limits. With an increased number of groups, MG-CFA becomes increasing impractical because it requires exact equivalence, that is all parameter exactly equal across all items, and therefore, testing for equivalence is often neglected. The Alignment method is an alternative approach to MG-CFA. In the present article, we will explore the use of the Alignment method to make possible what has up to now been more difficult to achieve: to compare latent means across many groups.

This paper will first describe MG-CFA and contrast it to the Alignment method theoretically. Next, we will apply both methods to a set of data from eight countries across eight measures. We will use this example to reflect on similarities and differences between both methods and their results. Finally, we will inspect latent mean differences across countries based on both methods and compare the pattern of these differences to previous research in order to contextualize the meaning of our results in the framework of the conceptual knowledge of the variables at hand.

Measurement Invariance: A Primer

Establishing full measurement invariance of a psychological self-report measure (a "scale") validates the comparisons of groups by insuring that the central parameters defining a measurement model – that is, the factor structure (i.e. number of factors), loadings (i.e. unit of a scale) or intercepts (i.e. original of the scale) – do not systematically depend on irrelevant factors, such as group, culture, or measurement occasions. Lack of MI can lead to erroneous conclusions, in that the observed mean differences are not due to latent mean differences but rather due to irrelevant differences between groups, such as response sets (Chen, 2008; van de Vijver & Tanzer, 2004). Some reasons for the lack of MI include sampling, translation issues, different interpretation of the questions, and cultural difference in social desirability and acquiescence (Heath, Martin & Sprekesen, 2009; van de Vijver & Tanzer, 2004).

Given its importance for valid research practice, it is surprising how often MI is not tested in published research. In a review of cross-cultural research between 2008 and 2015 only 17% of articles published in the *Journal of Cross Cultural Psychology* and only 1% of articles published in *Personality and Social Psychology Bulletin* assessed measurement equivalence (Boer, Hanke, & He, 2018). Chen (2008) also found only 17% of cross-cultural research studies between 1985 and 2005 in the *Journal of Personality and Social Psychology* tested measurement invariance. While some scholars have pointed to an inadequate education of the importance and execution of invariance testing (Steenkamp & Baumgartner, 1998), one reason why measurement invariance is so often disregarded in cross-cultural research is that the methods used to scrutinize MI may often be difficult to apply or only feasible under certain circumstances. For instance, the traditional method of multiple-group confirmatory factor analysis (MG-CFA) is often unattainable for many groups (Byrne & van de Vijver 2017; Kim, Cao, Wang & Nguyen, 2017; Marsh et al., 2016; Rutkowski & Svetina 2014). While the literature is not precise at what is meant by many groups, finding invariance becomes more difficult as the number of groups increases, as does the number of possible violations: in MG-CFA, the statistical null hypothesis of measurement invariance is defined as

exact equivalence of all parameters (i.e., zero deviation) across groups. This null hypothesis might be too strict (for a comparable argument, see Cohen, 1994), so MG-CFA is likely to reject models that are "practically comparable" if the deviation of parameters across groups is non-zero (Lomazzi, 2018). When researchers have unique, many-group data sets, and want to publish their findings, exact measurement equivalence practices are frequently impossible and therefore prohibitive to doing a comparative analysis of groups. This dilemma was described by Byrne and van de Vijver (2017): "As a result, these limitations have remained a major impediment to advancing our substantive knowledge of cross-group differences within the context of a wide variety of disciplines and in the conduct of numerous important large scale studies both nationally and cross-nationally" (p. 540). Therefore, establishing the utility of new methods is of utmost importance.

Multiple-group confirmatory analysis is first attributed to the seminal work of Jöreskog (1971) and later extended to test invariance of latent factor means (Sörbom, 1974). The MG-CFA method for MI is tested in steps: after assessing model fit for each group separately, model fit is established across groups, that is, *configural invariance* or *factor form invariance*, which, if found, shows that the same items are associated with the same latent factor. Next, *metric* or *"weak" factorial invariance* is tested, which assumes that the items have similar factor loadings across groups, which in turn allows for predictive relationships to be compared across groups. This is followed by *scalar* or *"strong" invariance* which, when established, means the item intercepts are equivalent across groups and is sufficient for factor mean comparisons. Following this is *residual invariance* or *"strict" invariance* which compares the equivalence of construct variance; however, this level is rarely tested because it is broadly agreed upon as not necessary for latent mean comparison (Cheung, & Rensvold, 2000). At each step, the more restrictive model is compared to the previous, lesser restrictive model, and when a significantly worse fit is found based on previously established benchmarks, then invariance is rejected (see Meredith, 1993; Vandenberg & Lance, 2000).

Traditionally, the changes in the chi square statistic were used as an indicator; however, because chi square is found to be highly sensitive to sample size, it is recommended to use the change in CFI of -.01(ΔRMSEA <0.015; ΔSRMR <0.03, Cheung & Rensvold, 2002; Chen, 2007; Meade et al., 2008). However, with the issues of multiple groups, some further relaxing for the restriction for scalar invariance has been recommended (Kim et al., 2017; Rutkowski & Svetina, 2014).

While MG-CFA is the most common method for testing MI, it has been most widely used in the context of two groups (Kim et al., 2017; Rutkowski & Svetina, 2014). There is an ongoing discussion in the literature and cross-cultural research field about the issues around how to handle a comparison of more than two to three groups. MG-CFA is now recognized as too stringent for numerous groups (in the context of large-scale cross-national studies, this usually refers to twenty or more groups, Byrne & van de Vijver 2017; Kim et al., 2017; Marsh et al.; 2016, Rutkowski & Svetina 2014). However, as it is the most normative standard for MI testing, it is a good benchmark to compare other methods.

As mentioned above, the MG-CFA method is found to be impractical for a large number of groups because exact invariance is often not achieved in practice. Even though this is widely acknowledged as an issue with large scale projects (20+ groups), it is also a problem with small and midsized numbers of different groups, which means three or more groups (Asparouhov & Muthén, 2014; Marsh et al., 2016; van de Schoot, Kluytmans, Tummers, Lugtig, Hox, & Muthen, 2013). Particularly, in a comparison between 3 to 10 countries (or heterogeneous groups), the literature is vaguer as to the utility of alternative methods for testing measurement invariance.

When a research question depends on group mean comparisons, what options are there when one fails to find strict MI? Typical solutions are to exclude items or groups, limiting the analysis to subgroups of countries, and partial invariance (explained below). The former is not ideal, firstly, because by deleting items, one sacrifices content validity. Secondly, reducing

the number of countries or the breadth of comparison dimensions excludes what might be vital information to the hypothesis at hand (Lomazzi, 2018). This is especially drastic with midsized number of groups (see Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014, for further stipulated on the possibilities with non-invariant data).

Partial invariance occurs when only a subset of parameters is invariant, while another subset varies across groups. In Boer et al.'s (2017) review of literature in *Journal of Cross Cultural Psychology* mentioned above, out of the 64 studies in which the equivalence of the measurement model has been tested, 46 reported some form of partial invariances. The literature says that just one item with non-invariant intercept can lead to erroneous conclusions when observed mean differences (i.e. differences between groups on an aggregate measure) are interpreted (Meredith, 1993; Steinmetz, 2013). However, if item parameters are invariant for at least 50% of the items in the scale, one can compare latent means, that is, estimated means of the latent factor on which the items load (Byrne et al., 1989, Reise, Widaman, & Pugh, 1993, Steenkamp & Baumgartner, 1998, Steinmetz, 2013). Although this is the implication of some simulation studies that investigate the consequences of partial MI, there is also debate about the soundness of the stepwise strategy (explained below) and conflicting results in simulation studies (Marsh et al., 2016; Rensvold & Cheung, 2001; Vandenberg, 2002; van de Schoot et al., 2013). Partial invariance stepwise procedure has an inherent risk of producing a model that deviates substantially from the true model because it depends on the post-hoc freeing of parameters. This added element of chance means the results might not be replicable (Lomazzi, 2018, Marsh et al., 2016; MacCallum, Roznowski & Necowitz, 1992). As Marsh et al. (2016) explain, "…there is no guarantee that the stepwise selection process, based on freeing factor loadings and intercepts to achieve an acceptable goodness of fit, will facilitate this objective of providing unbiased means" (p. 12).

Due to the limitation of MG-CFA for a large number of groups, researchers have explored other methods, including multilevel confirmatory factor analysis (Davidov, Dülmer,

Schlüter, Schmidt, & Meuleman, 2012), multilevel factor mixture modeling, ESEM (Byrne & van de Vijver, 2017), Bayesian approximation MI (Muthén & Asparouhov, 2013a) and Alignment optimization method (Asparouhov & Muthén, 2014). Based on a review of these methods by Kim et al. (2017), Alignment method was chosen as the best suited to be an alternative to look at MI in a mid-range number of groups.

Alignment Optimization

Alignment optimization was first introduced by Asparouhov and Muthén (2014) as an alternative to MG-CFA. The idea is that estimates of group factor means can be made without requiring exact MI; therefore, Alignment is a method of approximate rather than exact invariance. It is theorized that by taking some amount of non-invariance into account, Alignment optimization provides more reliable models (Byrne & van de Vijver, 2017; Marsh et al., 2016). Furthermore, in contrast to MG-CFA, Alignment starts with a common configural model that contains all group for which the intercepts and loadings are unconstrained rather than a baseline model for each group separately as with MG-CFA (Byrne & van de Vijver, 2017). From the configural model the process uses maximum likelihood (ML) to estimate an optimal set of measurement parameters and computes approximations based on that. This optimization procedure involves applying a simplicity function that works similar to the rotation criteria in EFA; that is, it simplifies the interpretation while the fit is the same as the configural model, i.e. the unrotated model. This allows for differences in model parameter estimates by using the most relaxed, thus best fit model. Means and variances of the latent variable can be computed on the basis of the optimal model.

The execution is a two-step process: first, a configural model that represents the best fitting model among all groups is established when fixing the factor means to 0 and fixing variances to 1, without constraining the loadings and intercepts. Rather than assuming invariance, Alignment operates with the assumption that there is a degree of non-invariance and the goal is to keep it to a minimum. In the next step, the factor means and variances are

freely estimated and undergo an optimization process for every group factor mean and item parameters (see Asparouhov & Muthén, 2014). Once the minimization point has been reached, the analysis compares factor means and factor variances across groups using a "post-estimation algorithm" and identifies each model parameter (i.e., loadings and intercepts) that is significantly different from the average of that parameter across all groups. This is done with multiple pairwise comparisons (adjusted for alpha inflation). If estimates are not significantly different they are connected creating an invariant set to which subsequent groups' parameters are compared to, while those that are significantly different are flagged as noninvariant. The output provides the latent means based on this model as well as the parameters flagged as non-invariant (see Asparouhov & Muthén, 2014, for a detailed description of the statistical equations for the process).

To our knowledge, there is only one simulation study that compares partial invariance with MG-CFA to Alignment. Marsh et al., (2016) found that the alignment method shows less bias for both scalar and partial scalar models. Our goal is to follow up this work with real data that compares the results of group-level latent mean comparisons under partial invariance found using MG-CFA or Alignment methods.

The Present Research

In the present study we want to investigate the implications of using traditional MG-CFA in comparison to Alignment optimization for testing mean differences between samples from different countries for eight measures. Up till now, research on the difference between MG-CFA and Alignment has been mainly based on simulation studies or situations in which the number of groups was very large and MI testing using MG-CFA was feasible that is, the analysis showed too many mis-specified parameters that need to be adjusted so that partial invariance was also not feasible (Byrne & van de Vijver, 2017). Therefore, this study looks at a medium number of countries for which partial invariance is a feasible alternative and compares that to the analysis derived from the Alignment method in order to compare the two

methods in this context. To illustrate the resulting differences in implications, we also distinctly relate the results of the two methods to prior research and theorizing in order to scrutinize the validity of both methods.

Country-level variables

The goal of establishing measurement invariance is often to be able to compare and contrast the mean differences. Sometimes this is the goal of the research in itself, as in many large-scale research projects such as PISA (Program for the International Assessment of Student Achievement) in which countries are being compared (i.e., ranked) regarding education-oriented variables (i.e. Byrne & van de Vijver, 2017, Marsh et al., 2016). In other cases, the primary goal of the research is to foster our theoretical understanding of a psychological measure or to assist in the appropriate sample selection for a specific research question.

Therefore, in our comparison of methods, we will also look at contexts in which the variables measured here have been studied at a cultural level in prior studies to see where our results are comparable with the results of these prior studies and whether the degree of comparability depends on whether MI has been addressed with the traditional MG-CFA or the Alignment approach. Horizontal/Vertical Individualism/Collectivism (Triandis & Gelfand, 1998) are highly relevant constructs in cross-cultural research, so, although there are not many large scale cross-cultural studies with these variables, there is a theoretical framework that can assist with the interpretation of our results. Not all of the variables measured here have been studied extensively on a cultural level. Social Dominance Ordination (Pratto, Sidanius, Stallworth, & Malle, 1994) and Right Wing Authoritarianism (Altemeyer, 1981) have not explicitly been tested at the country level; however, past research shows their relationship to other variables that are measured at the country level such as economic equality and religiosity. Victim Sensitivity and Observer Sensitivity (Schmitt, Gollwitzer, Maes, & Arbach, 2005) have very little past research comparing countries; therefore, our result will be

a unique contribution. We take these previously developed variables and look at the relationship to other relevant cultural level variables to both compare the methods and to conceptually validate our findings.

Methods

Samples

The data for this study is the secondary analysis on materials collected for an independent project. This multinational project involved researchers in eight countries that were all tasked with finding participants in their respective countries. The aim was 200 participants per country; however, final participant numbers were varied due to the individual situation of the data collection location. In the case of the German sample, it was not possible to get more participants in the time frame allotted because recruitment depended on the size on the student sample pool at the time of data collection. The sample sizes, age range, and gender distribution can be seen in Table 1. All research groups were instructed to follow the translation-back translation process from originally English materials (preregistration document available at: https://doi.org/10.17605/OSF.IO/8MZCN). All participants completed the material on an online platform. The sample from Germany was collected by the first author. Participants were psychology Bachelor students who received course credits for their participation.

Table V-1

Sample Information

Country	Sample size	Age mean	age SD	% female
1. USA	N= 130	26,01	5,40	53,4
2. UK	N = 294	19,13	2,61	84,7
3. Canada	N = 72	25,96	4,83	40,8
4. Germany	N = 116	21,75	3,80	75,9
5. India	N = 177	20,90	4,59	52,0
6. Malaysia	N = 169	19,35	2,20	47,6
7. Nigeria	N = 205	21,92	3,35	49,0
8. Peru	N = 91	23,61	4,27	51,1

Measures

The measures we analyzed for the present study are the following eight scales:

Horizontal/Vertical Individualism/Collectivism Scale – 6-point Likert Scale (1 = Strongly Disagree; 6 = Strongly Agree) 4 items each. *Horizontal Collectivism (HC)* measures the degree to which people see themselves as similar to other people and emphasize common goals as well as interdependence (e.g. "To me, pleasure is spending time with others." *Vertical Collectivism (VC)* measures the degree to which people emphasize the importance of the group and are willing to go along with the authority of the group even when it does not fit one's own interests (e.g., "Family members should stick together no matter what sacrifices are required.") *Horizontal Individualism (HI)* measures the degree to which people value both equality and freedom and want to be unique (e.g., "I'd rather depend on myself than others.") *Vertical Individualism (VI)* measures the degree to which people want to be distinguished in order to gain status and value freedom over equality (e.g., "Competition is the law of nature," Triandis & Gelfand, 1998).

Social Dominance Ordination SDO-6 Scale (Hierarchy Items) – 7-point Likert Scale (1 = Strongly Disagree; 7 = Strongly Agree), 8 items. SDO measures one's motivation for social hierarchy and group based dominance (e.g., "It is probably a good thing that certain groups are at the top and other groups are at the bottom," Pratto et al., 2000).

Right Wing Authoritarianism (RWA) short scale – 7-point Likert Scale (1 = Strongly Disagree; 7 = Strongly Agree), 7 items. RWA measures the belief in the uncritical subjection to authority and defense of authoritative norms (e.g., "Our country desperately needs a mighty leader who will do what has to be done to destroy the radical new ways and sinfulness that are ruining us;" Manganelli, Rattazzi, Bobbio, & Canova, 2007). See Tables 2a-h for the correlation between the variables for each country.

Victim Sensitivity & Observer Sensitivity (Justice Sensitivity subscales) – 6-point Likert Scale (1 = Strongly Disagree; 6 = Strongly Agree), 10 items each. *Victim Sensitivity*

(VS) measures sensitivity to being the victim of injustice, (e.g., "It bothers me when others receive something that ought to be mine"). *Observer Sensitivity (OS)* measures sensitivity to injustice toward others (e.g., "I am upset when someone does not get a reward he/she has earned," Schmitt et al., 2005).

Table V-2

Correlation between measured variables for each country

Table V-2a.

USA N= 130

	VS	OS	HC	VC	HI	VI	SDO	RWA
VS	1							
OS	,505**	1						
HC	-,188*	,230**	1					
VC	-0,042	0,156	,402**	1				
HI	0,143	0,049	-0,066	-0,109	1			
VI	,344**	0,074	-0,085	0,069	0,148	1		
SDO	,177*	-0,043	0,001	,212*	-0,029	,429**	1	
RWA	0,125	-0,067	-0,021	,362**	-0,033	,337**	,747**	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).
VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Table V -2b.

UK N = 294

	VS	OS	HC	VC	HI	VI	SDO	RWA
VS	1							
OS	,394**	1						
HC	-,148*	,324**	1					
VC	0,017	,224**	,445**	1				
HI	0,065	,116*	0,031	-0,055	1			
VI	,316**	-0,029	-,191**	-0,079	,236**	1		
SDO	,194**	-,144*	-,218**	-,143*	-,118*	,269**	1	
RWA	0,001	-,135*	0,041	,204**	0	,217**	,330**	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).
VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Table V -2c.

Canada N = 72

	VS	OS	HC	VC	HI	VI	SDO	RWA
VS	1							
OS	,393**	1						
HC	0,139	,539**	1					
VC	0,213	,263*	,497**	1				
HI	0,114	-0,084	0,005	-0,037	1			
VI	,496**	0,203	-0,009	,234*	0,16	1		
SDO	0,128	-0,057	-0,129	0,045	0,024	,396**	1	
RWA	0,13	-0,13	0,097	,299*	0,159	,323**	,597**	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).
VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Table V -2d.

Germany N = 116

	VS	OS	HC	VC	HI	VI	SDO	RWA
VS	1							
OS	,410**	1						
HC	,236*	,498**	1					
VC	,244**	,313**	,494**	1				
HI	,195*	0,179	-0,077	0,04	1			
VI	,248**	-0,005	-0,025	0,18	,303**	1		
SDO	0,167	-0,152	-,193*	0,082	0,068	,396**	1	
RWA	0,015	-0,087	-0,048	,277**	0,077	,371**	,565**	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).
VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Table V -2e.

India N = 177

	VS	OS	HC	VC	HI	VI	SDO	RWA
VS	1							
OS	,456**	1						
HC	0,05	,380**	1					
VC	,172*	,312**	,544**	1				
HI	,226**	0,132	,367**	,273**	1			
VI	,449**	0,136	0,019	,215**	,196**	1		
SDO	,160*	-0,031	-0,142	-0,067	-0,118	,355**	1	
RWA	,242**	,235**	,230**	,393**	,204**	,284**	,380**	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).
VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Table V -2f.

Malaysia N = 169

	VS	OS	HC	VC	HI	VI	SDO	RWA
VS	1							
OS	,328**	1						
HC	-0,059	,512**	1					
VC	0,079	,371**	,650**	1				
HI	,206**	0,147	0,126	,171*	1			
VI	,426**	-0,072	-0,147	0,004	,266**	1		
SDO	,179*	-,306**	-,275**	-0,133	0,012	,369**	1	
RWA	0,079	0,101	,270**	,326**	,227**	0,128	,279**	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).
VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Table V -2g.

Nigeria N = 205

	VS	OS	HC	VC	HI	VI	SDO	RWA
VS	1							
OS	,594**	1						
HC	0,093	,193**	1					
VC	,224**	,334**	,551**	1				
HI	,229**	,272**	,322**	,351**	1			
VI	,536**	,387**	,168*	,367**	,402**	1		
SDO	,142*	0,05	0,002	0,096	0,055	,246**	1	
RWA	,158*	,201**	,381**	,474**	,361**	,254**	0,036	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).
VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Table V -2h.

Peru N = 91

	VS	OS	HC	VC	HI	VI	SDO	RWA
VS	1							
OS	,643**	1						
HC	-,227*	0,01	1					
VC	0,049	0,102	0,202	1				
HI	,214*	0,157	-0,133	-,228*	1			
VI	,426**	,236*	-,224*	,367**	0,095	1		
SDO	0,133	-0,068	-0,043	,214*	-0,062	,339**	1	
RWA	-0,037	-0,109	0,027	,387**	-0,15	,371**	,536**	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).
VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Analysis Strategy

MG-CFA. For the multiple group confirmatory factor analysis, we followed a stepwise approach of establishing measurement invariance for each of the scales (see Meredith, 1993; Vandenberg & Lance, 2000). We completed this analysis with lavaan (Rosseel, 2012) and semTools for R (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2018). We looked at the fit indices that are usually reported in the literature and used values for acceptable model fit that have been suggested by Marsh, Hau, & Wen, (2004) : root mean square error of approximation (RMSEA; ≤.080), standardized root mean square residual (SRMR; ≤.080), Tucker–Lewis index (TLI, >.90), and comparative fit index (CFI; >.90). However, these are recommendations for model fit rather than a golden rule; therefore, it is important that we compare the relative fit at each level of invariance (Marsh et al., 2016). Here we focused on a change of CFI of < .01 to represent acceptable difference in fit.

If scalar invariance was not achieved, we proceeded to test for partial invariance. For partial invariance, we followed the stepwise approach by examining the mis-specified parameter, freeing one parameter at a time, starting with the largest value (Byrne, Shavelson, & Muthén, 1989; Steenkamp & Baumgartner, 1998). For each parameter freed, we again checked the comparative fit. We repeated this process until half the parameters of the items were freed. If partial invariance was still not established, we began the process again by excluding countries with high modification indices or high chi-square values in the strict invariance model. Once we established partial invariance at the scalar level for all measures, we used these models to calculate the latent means and compared them with Germany as the reference group. We used Germany as the reference group because this was the sample collected by the primary author and where some of the scales were developed (i.e. VS and OS).

Alignment. Using Mplus version 8.1 (Muthén & Muthén, 2018) and the syntax in the Appendix, we followed guidelines from Asparouhov and Muthén (2014) for the Alignment

optimization, imputing each measure in a separate analysis. We used FIXED alignment rather than FREE as recommended by Marsh et al., (2016), Kim et al., (2017), and Jang et al., (2017). The FIXED option fixes the latent means of one group, in this case Germany to 0 and variance to 1, thus serving as the reference group (see Asparouhov and Muthén, 2014, for a detailed discussion on the difference between FIXED and FREE options).

Research has been mixed on the degree of acceptable non-invariant parameters with the Alignment method that still allows a valid comparison. In the original simulation study for the Alignment method, Asparouhov and Muthén (2014) mentioned 20% non-invariant parameters as acceptable; later, Muthén and Asparouhov (2014) adjusted to 25% as acceptable. Flake & McCoach (2018) stipulated there should be less than 29% non-invariant parameters. However, as this is a relatively new field, Muthén and Asparouhov (2014) recommended when above 25%, a Monte Carlo simulation should supplement the judgement of acceptability. Following this recommendation, Lomazzi (2018) found 21% acceptable. However, this is an ongoing open question for the Alignment method.

Comparison

Using the models resulting from the invariance testing by both methods, we compared the composite means and latent means. Furthermore, we related our findings to past research that looked at the variables of this study at a country level and we investigated their relation with other theoretically related variables. The validation criterion for each scale is based on previous theoretical knowledge of the variable and the predictions outlined below.

HC/VC/HI/VI. These scales have been used the most in cross-cultural research; however, there are few large-scale studies that compare means across the relevant countries. Horizontal/vertical individualism and collectivism are thought to be individual level variables (Triandis, 2001; Markus & Kitayama, 1991); therefore, there are not many large scales studies that compare these scales at a country level. However, the broad understanding is that these individual level variables correspond to country level difference. Hofstede (1980, 2010) has

looked extensively at country level variables. Therefore, we look at the relationship between HC/VC/HI/VI and both Hofstede's individualism-collectivism (IND-COL) and power distance. IND-COL is predicted to be positively correlated with HI/VI and negatively correlated with HC/VC. Power distance is expected to reflect the hierarchy versus egalitarian aspects of the construct in that power distance is positively correlated with VC/VI and negatively correlated with HC/HI. We used the values found in Hofstede, Hofstede, & Minkov, (2010), which include country level values for all of our countries.

SDO/RWA. In a review of the literature, there are some studies that have looked at the relationship between SDO and RWA across cultures and found they are related but distinct concepts especially in regard to the influence on prejudice (Duckitt, & Sibley, 2007; Perry, Sibley, & Duckitt, 2013). Therefore, we expect them to be correlated on the country level as well. For an overall theory of country-level differences, Fischer, Hanke, Sibley, (2012) completed a meta-analysis that measured SDO, looking at other country level variables in relation to the calculated country level means. They found a correlation between SDO and country level variables of gender equality, democracy, Gross National Income (GNI), and egalitarian values. We looked at the latter two, first with the GNI index including both the GNI country rank and national income expressed as purchase parity measures value in dollars (World Bank, 2017). This was followed by the same country-level scores from the Schwartz Value Survey of the egalitarian scale of Fischer et al., (2012). Participants rated the importance of the value responding to "as a guiding principle in MY life" on a 9-point scale. Schwartz (1994, 2006) explains this response format and sample information. We used the average scores of the student and teacher data. We predict that SDO will be significantly negatively correlated with Gross National Income (GNI) and egalitarian values.

With the measure of RWA, few studies have collected data in non-western countries (i.e,. meta-analyses, Sibley, & Duckitt, 2008; Perry, et al., 2013). However, some research shows a positive relationship between RWA and religious fundamentalism (Altemeyer, &

Hunsberger, 1992); therefore, we looked at country level religiosity. To do this, we took the variable from the World Value Survey that measured "importance of religion" from 1 = "very important"; 4 = "not at all important" (reverse coded). To have data from all the countries we have in our data, we had to look at both wave 5 (2005-2009) and wave 6 (2010-2014, Inglehart et al., 2014a, 2014b). Data for Canada and UK were from wave 5 and other countries from wave 6. Sample size for the countries in the present analysis was between 1026 (in Canada) and 4019 (in India). We predict RWA to be significantly positively correlated with religiosity. Furthermore, we predict our measure of SDO and RWA to be positively correlated.

VS/OS. These two scales are mostly absent in previous cross-cultural research or theory of culture level differences. Few papers have looked at culture and justice sensitivity scales. In Magraw-Mickelson and Gollwitzer (2018), VS was measured in Germany, Japan and the US; however, there was no comparison of group level means due to the lack of scalar invariance. In another comparison, Wu and colleagues (2014) looked at observer sensitivity (and beneficiary sensitivity) in China, Germany, and Russia. They found that Chinese participants scored lower on OS compared to German participants, whereas Chinese were not significantly different from Russian participants. Furthermore, OS was not related to the cultural variables individualism and collectivism. According to the literature, we would expect similar findings regarding VS.

Results

The results of the MG-CFA invariance testing can be seen in Tables 3-7. For each measure, first comes the single group solution and initial test of full invariance followed by the results for partial invariance. The results show that we found metric invariance[13] for all measures but were unable to find scalar invariance. However, we did find partial scalar

[13] For the SDO scale the CFI difference was -.0104, borderline acceptable value.

invariance with HC/VC/HI/VI and SDO; the necessary intercepts freed for the partial model are listed in the Table. For VS, OS, and RWA, partial scalar invariance was only found for a subset of countries (also listed in Tables 3-7).

For the Alignment optimization, we first checked the overall degree of non-invariance for each scale. As mentioned above, the existing guidelines are that 25% of non- invariance is problematic or more strictly that those above 20% should require a closer look. None of our scales came close to this number. The highest degree of non-invariance across parameters was 2% of intercepts for the SDO scale. Therefore, we concluded that Alignment optimization is feasible for all measures (see Tables 8-12).

Comparison of mean differences

Table 13 shows a comparison of the means and the significance of mean differences. First, listed in the table are the simple composite means for each measure for each country followed by the latent means found with MG-CFA scalar partial invariance and finally the latent means from the alignment optimization. Because scalar invariance was not established for any of the scales, the composite scores provide an invalid comparison (Meredith, 1993; Steinmetz, 2013). However, these means provide reference for the degree to which the composite means may be inaccurate. Although there were some shifts in the rank order of the countries between the composite mean, MG-CFA latent means and Alignment latent means, those position shifts were not significant. For example, the composite mean of the VS scale is lower in Peru than in the USA and in Canada, but the latent mean (estimated on the basis of the Alignment procedure) is higher in Peru than in the USA or in Canada. However, in all analyses the difference between these three countries is insignificant. However, there were numerous discrepancies in whether the mean differences between countries reached significance. For example, the composite mean of the VS scale for Nigeria was significantly greater than that of India and Germany; however, in the latent mean comparison (estimated on the basis of the Alignment procedure), although Nigeria was still ranked higher, the value was

not significantly greater compared to India or Germany. This is true for both the comparison of the composite means and latent means, and between the latent means achieved by different methods. Overall, there are more significant differences found when looking at the raw composite score compared to the latent means and in the MG-CFA latent means compared to the Alignment latent means, which points to a systematic directionality of the bias in the raw composite scores, in that mean differences is inflated the between countries. See Table 13 for all comparisons and differences in significance.

Table V-3

Measurement invariance testing for Collectivism-Individualism

Single Group Solutions	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Canada (n=67)	177,76	98					.109 (.083 - .135)	0,109	0,826	0,787
Germany (n=114)	155,61	98					.071 (.049 - .092)	0,085	0,922	0,905
India (n=169)	191,62	98					.078 (.061 - .094)	0,094	0,894	0,87
Malaysia (n=160)	198,58	98					.081 (.065 - .098)	0,084	0,892	0,868
Nigeria (n=187)	234,26	98					.086 (.072 - .101)	0,080	0,862	0,831
Peru (n=88)	170,37	98					.091 (.068 - .114)	0,110	0,861	0,830
UK (n=290)	299,43	98					.085 (.074 - .096)	0,070	0,878	0,851
USA (n = 129)	228,26	98					.103 (.085 -.120)	0,087	0,846	0,812
Measurement Invariance	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA**	**SRMR**	**CFI**	**TLI**
Equal form	1655,90	784					0,087	0,081	0,876	0,848
Equal factor loadings	1808,30	868	2,08	152,40	84	0,00	0,086	0,089	0,866	0,852
Equal indicator intercepts	2207,10	952	2,32	398,80	84	0,00	0,094	0,097	0,821	0,82
Partial Invariance 'HI1~1', 'HI2~1', 'VI4~1', 'VC1~1', 'HC1~1', 'VC2~1', 'VI1~1'	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA**	**SRMR**	**CFI**	**TLI**
Equal form	1655,90	784					0,087	0,081	0,876	0,848
Equal factor loadings	1808,25	868	2,08	152,35	84	0,00	0,086	0,089	0,866	0,852
Equal indicator intercepts	1912,03	903	2,12	103,78	35	0,00	0,087	0,091	0,856	0,847

Table V-4

Measurement invariance testing for SDO

Single Group Solutions	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Canada (n=67)	48,18	20					.140 (.090 - .191)	0,055	0,931	0,904
Germany (n=114)	47,49	20					.109 (.069 - .149)	0,057	0,941	0,918
India (n=169)	96,837	20					.147 (.119 -.177)	0,060	0,877	0,828
Malaysia (n=160)	33,70	20					.064 (.021 - .100)	0,052	0,981	0,973
Nigeria (n=187)	67,84	20					.108 (.080 -.137)	0,060	0,881	0,833
Peru (n=88)	56,56	20					.141 (.098 - .185)	0,081	0,858	0,801
UK (n=290)	157,69	20					.153 (.131 - .176)	0,067	0,87	0,818
USA (n = 129)	119,82	20					.195 (.163 - .230)	0,060	0,898	0,858
Measurement Invariance	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA**	**SRMR**	**CFI**	**TLI**
Equal form	628,11	160					0,137	0,054	0,905	0,867
Equal factor loadings	728,24	186	3,92	100,13	26	0,000	0,126	0,087	0,894	0,887
Equal indicator intercepts	1165,34	258	4,52	437,10	72	0,000	0,15	0,112	0,815	0,840
Measurement Invariance (-USA, -Nigeria, -Peru; SDO8 ~1, SDO5 ~1, SDO4 ~1, SDO7 ~1)	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA**	**SRMR**	**CFI**	**TLI**
Equal form	383,89	100					0,131	0,052	0,913	0,878
Equal factor loadings	428,83	128	3,35	44,94	28	0,022	0,119	0,077	0,908	0,899
Equal indicator intercepts	467,73	140	3,34	38,90	12	0,000	0,119	0,080	0,900	0,900

Table V-5

Measurement invariance testing for RWA

Single Group Solutions	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Canada (n=67)	30,31	14					.127 (.064 - .190)	0,056	0,929	0,893
Germany (n=114)	22,60	14					.073 (.000 - .126)	0,043	0,972	0,959
India (n=169)	39,05	14					.101 (.064 - .139)	0,050	0,94	0,91
Malaysia (n=160)	14,47	14					.014 (.000 - .076)	0,035	0,999	0,998
Nigeria (n=187)	16,69	14					.031 (.000 - .077)	0,043	0,989	0,984
Peru (n=88)	39,33	14					.140 (.089 - .193)	0,056	0,925	0,887
UK (n=290)	50,65	14					.094 (.067 - .123)	0,043	0,95	0,926
USA (n = 129)	24,73	14					.077 (.019 - .125)	0,027	0,984	0,977
Measurement Invariance	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA**	**SRMR**	**CFI**	**TLI**
Equal form	237,82	112					0,085	0,038	0,962	0,943
Equal factor loadings	308,74	154	2,00	70,91	42	0,003	0,080	0,071	0,953	0,949
Equal indicator intercepts	591,94	196	3,02	283,20	42	0,000	0,113	0,096	0,881	0,898
Measurement Invariance (- India and Malaysia 'RWA6 ~1', 'RWA1 ~1','RWA4 ~1','RWA2 ~1')	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA**	**SRMR**	**CFI**	**TLI**
Equal form	184,31	84					0,089	0,038	0,961	0,941
Equal factor loadings	235,49	114	2,07	51,18	30	0,009	0,084	0,070	0,952	0,947
Equal indicator intercepts	267,79	124	2,16	32,30	10	0,000	0,087	0,073	0,944	0,942

Table V-6

Measurement invariance testing for Victim Sensitivity

Single Group Solutions	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Canada (n=69)	69,62	35					.120 (.078 - .161)	0,078	0,849	0,806
Germany (n=114)	47,66	35					.056 (.000- .093)	0,054	0,961	0,949
India (n=173)	120,21	35					.119 (.096 - .142)	0,073	0,826	0,777
Malaysia (n=162)	76,14	35					.085 (.059 - .111)	0,050	0,939	0,921
Nigeria (n=188)	132,11	35					.121 (.100 - .144)	0,079	0,843	0,798
Peru (n=88)	80,91	35					.122 (.087 - .157)	0,084	0,838	0,792
UK (n=282)	122,43	35					.094 (.076 - .113)	0,055	0,903	0,875
USA (n = 129)	97,48	35					.118 (.090 - .146)	0,065	0,898	0,869
Measurement Invariance	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Equal form	746,55	280					0,105	0,060	0,887	0,855
Equal factor loadings	831,44	343	2,42	84,89	63	0,034	0,097	0,080	0,882	0,876
Equal indicator intercepts	1104,34	406	2,72	272,90	63	0,000	0,107	0,095	0,831	0,850
Partial Invariance (-India; 'VS9 ~1', 'VS7 ~1', 'VS6 ~1', 'VS8 ~1', 'VS3 ~1')	**χ2**	**df**		**χ2diff**	**Δdf**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Equal form	626,35	245					0,103	0,059	0,895	0,865
Equal factor loadings	704,88	299	2,4	78,53	54	0,02	0,096	0,080	0,888	0,882
Equal indicator intercepts	764,74	323	2,4	59,87	24	0,00	0,096	0,084	0,879	0,882

Table V-7

Measurement invariance testing for Observer Sensitivity

Single Group Solutions	**χ2**	**df**		**χ2diff**	**∆df**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Canada (n=67)	98,14	35					.164 (.126 - .203)	0,096	0,766	0,699
Germany (n=114)	92,36	35					.120 (.090 - .150)	0,066	0,892	0,861
India (n=169)	150,70	35					.140 (.117 - .163)	0,092	0,753	0,683
Malaysia (n=160)	110,51	35					.116 (.092 - .141)	0,058	0,907	0,88
Nigeria (n=187)	180,72	35					.149 (.128 - .171)	0,088	0,768	0,702
Peru (n=88)	73,85	35					.122 (.076 - .148)	0,070	0,877	0,842
UK (n=290)	206,21	35					.130 (.113 - .147)	0,077	0,842	0,797
USA (n = 129)	104,93	35					.124 (.097 - .152)	0,070	0,898	0,868
Measurement Invariance	**χ2**	**df**		**χ2diff**	**∆df**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Equal form	1017,40	280					0,132	0,071	0,846	0,802
Equal factor loadings	1069,00	343	3,12	51,60	63,00	0,847	0,119	0,082	0,848	0,841
Equal indicator intercepts	1279,80	406	3,15	210,80	63,00	0,000	0,120	0,092	0,818	0,838
Partial Invariance (- Nigeria; 'OS6~1', 'OS1~1', 'OS3~1')	**χ2**	**df**		**χ2diff**	**∆df**	**sig**	**RMSEA (90% CI)**	**SRMR**	**CFI**	**TLI**
Equal form	836,70	245					0,129	0,069	0,858	0,817
Equal factor loadings	877,19	299	2,93	40,49	54,00	0,913	0,115	0,080	0,861	0,854
Equal indicator intercepts	950,95	335	2,84	73,76	36,00	0,000	0,112	0,084	0,852	0,861

Table V-8

Alignment measurement (non) invariance for intercepts and loadings for Horizontal/Vertical Individualism/Collectivism scales

Intercepts/Thresholds								
HC1	1	2	3	(4)	5	6	7	8
HC2	1	2	3	4	5	6	7	8
HC3	1	2	3	4	5	6	7	8
HC4	1	2	3	4	5	6	7	8
VC1	1	2	3	4	5	6	7	8
VC2	1	2	3	4	5	6	7	8
VC3	1	2	3	4	5	6	7	8
VC4	1	2	3	4	5	6	7	8
HI1	1	2	3	4	5	6	7	8
HI2	1	2	3	4	5	6	7	8
HI3	1	2	(3)	4	5	6	7	8
HI4	1	2	3	4	5	6	7	8
VI1	1	2	3	4	5	6	7	8
VI2	1	2	3	4	5	6	(7)	8
VI3	1	2	3	4	5	6	7	8
VI4	1	2	3	(4)	5	6	7	8
Loadings								
HC1	1	2	3	4	5	6	7	8
HC2	1	2	3	4	5	6	7	8
HC3	1	2	3	4	5	6	7	8
HC4	1	2	3	4	5	6	7	8
VC1	1	2	3	4	5	6	7	8
VC2	1	2	3	4	5	6	7	8
VC3	1	2	3	4	5	6	7	8
VC4	1	2	3	4	5	6	7	8
HI1	1	2	3	4	5	6	7	8
HI2	1	2	3	4	5	6	7	8
HI3	1	2	3	4	5	6	7	8
HI4	1	2	3	4	5	6	7	8
VI1	1	2	3	4	5	6	7	8
VI2	1	2	3	4	5	6	7	8
VI3	1	2	3	4	5	6	7	8
VI4	1	2	3	4	5	6	7	8

Table V-9

Alignment measurement (non) invariance for intercepts and loadings for Social Dominance Ordination scale

Intercepts/Thresholds								
SDO1	(1)	2	3	4	5	6	7	8
SDO2	1	2	3	4	5	6	7	8
SDO3	1	2	3	4	5	6	7	8
SDO4	1	2	3	(4)	5	6	7	8

SDO5	1	2	3	(4)	5	6	7	8
SDO6	1	2	3	4	5	6	(7)	8
SDO7	1	2	3	4	(5)	6	(7)	8
SDO8	1	2	3	(4)	5	6	7	8
Loadings								
SDO1	(1)	2	3	4	5	6	7	8
SDO2	1	2	3	4	5	6	7	8
SDO3	1	2	3	4	5	6	7	8
SDO4	1	2	3	4	5	6	7	8
SDO5	1	2	3	4	5	6	7	8
SDO6	1	2	3	4	5	6	7	8
SDO7	1	2	3	4	5	6	7	8
SDO8	1	2	3	4	5	6	7	8

Table V-10

Alignment measurement (non) invariance for intercepts and loadings for Right Wing Authoritarianism scale

Intercepts/Thresholds								
RWA1	1	2	3	4	5	6	7	8
RWA2	1	2	3	4	5	6	7	8
RWA3	1	2	3	4	5	6	7	8
RWA4	1	2	3	4	5	6	7	8
RWA5	1	2	3	4	5	6	7	8
RWA6	1	(2)	3	4	5	6	7	8
RWA7	1	2	3	4	5	6	7	8
Loadings								
RWA1	1	2	3	4	5	6	7	8
RWA2	1	2	3	4	5	6	7	8
RWA3	1	2	3	4	5	6	7	8
RWA4	1	2	3	4	5	6	7	8
RWA5	1	2	3	4	5	6	7	8
RWA6	1	2	3	4	5	6	7	8
RWA7	1	2	3	4	5	6	7	8

Table V-11

Alignment measurement (non) invariance for intercepts and loadings for Victim Sensitivity scale

Intercepts/Thresholds								
VS1	1	2	3	4	5	6	7	8
VS2	1	2	3	4	5	6	7	8
VS3	1	2	3	4	5	6	7	8
VS4	1	2	3	4	5	6	7	8
VS5	1	2	3	4	5	6	7	8
VS6	1	2	3	(4)	5	6	7	8
VS7	1	(2)	3	4	5	6	(7)	8

VS8	1	2	3	4	5	6	7	(8)
VS9	1	2	3	4	5	6	7	8
VS10	1	2	3	4	5	6	7	8
Loadings								
VS1	1	2	3	4	5	6	7	8
VS2	1	2	3	4	5	6	7	8
VS3	1	2	3	4	5	6	7	8
VS4	1	2	3	4	5	6	(7)	8
VS5	1	2	3	4	5	6	7	8
VS6	1	2	3	4	5	6	7	8
VS7	1	2	3	4	5	6	7	8
VS8	1	2	3	4	5	6	7	8
VS9	1	2	3	4	5	6	7	8
VS10	1	2	3	4	5	6	7	8

Table V-12

Alignment measurement (non) invariance for intercepts and loadings for Observer Sensitivity scale

Intercepts/Thresholds								
OS1	1	2	3	(4)	5	6	7	8
OS2	1	2	3	4	5	6	(7)	8
OS3	1	2	3	4	5	6	7	(8)
OS4	1	2	3	4	5	6	7	8
OS5	1	2	3	4	5	6	7	8
OS6	1	2	3	4	5	6	(7)	8
OS7	1	2	3	4	5	6	(7)	8
OS8	1	2	3	4	5	6	7	8
OS9	1	2	3	4	5	6	7	8
OS10	1	2	3	4	5	6	7	8
Loadings								
OS1	1	2	3	4	5	6	7	8
OS2	1	2	3	4	5	6	7	8
OS3	1	2	3	4	5	6	7	8
OS4	1	2	3	4	5	6	7	8
OS5	1	2	3	4	5	6	7	8
OS6	1	2	3	4	5	6	7	8
OS7	1	2	3	4	5	6	7	8
OS8	1	2	3	4	5	6	7	8
OS9	1	2	3	4	5	6	7	8
OS10	1	2	3	4	5	6	7	8

Table V-13

Mean comparisons 8 countries 8 scales

	RAW-Composite			MG-CFA			ALIGNMENT		
		Factor Mean	Groups With Significantly Smaller Factor Mean	HI1~1, HI2~1, VI4~1, VC1~1, HC1~1, VC2~1, VI1~1	Factor Mean	Groups With Significantly Smaller Factor Mean		Factor Mean	Groups With Significantly Smaller Factor Mean
Horizontal Collectivism	8. Peru	4,85	6, 5, 3, 1	7. Nigeria	0,119	3, 5, 2, 6, 1	8. Peru	0,172	6, 5, 3, 1
	7. Nigeria	4,77	6, 5, 3, 1	8. Peru	0,056	3, 5, 1	7. Nigeria	0,149	6, 5, 3, 1
	2. UK	4,68	1	4. Germany	0,000	3, 1	4. Germany	0,000	
	4. Germany	4,68		2. UK	-0,149	1	2. UK	-0,005	1
	6. Malaysia	4,59		6. Malaysia	-0,175	1	6. Malaysia	-0,147	
	5. India	4,55		5. India	-0,256		5. India	-0,188	
	3. Canada	4,48		3. Canada	-0,336		3. Canada	-0,221	
	1. USA	4,47		1. USA	-0,480		1. USA	-0,286	
Vertical Collectivism	7. Nigeria	4,96	6, 5, 8, 4, 3, 2, 1	7. Nigeria	0,940	4, 3, 5, 2, 6, 8,1	7. Nigeria	0,747	5, 2, 4, 1, 3, 8
	6. Malaysia	4,66	5, 8, 4, 3, 2, 1	6. Malaysia	0,466	4, 3, 2, 8, 1	6. Malaysia	0,464	2, 4, 1, 3, 8
	5. India	4,43	1, 8, 3	5. India	0,176	3, 1	5. India	0,244	1, 3, 8
	2. UK	4,30	1, 8, 3	4. Germany	0,000	3, 1	2. UK	0,032	1, 3, 8
	4. Germany	4,28	1, 8, 3	2. UK	-0,048	3, 1	4. Germany	0,000	1, 3, 8
	1. USA	3,97		8. Peru	-0,115	1	1. USA	-0,382	
	8. Peru	3,93		3. Canada	-0,504		8. Peru	-0,462	
	3. Canada	3,91		1. USA	-0,545		3. Canada	-0,533	
Horizontal Individualism	3. Canada	5,03	5, 8, 2, 6, 4	7. Nigeria	1,201	5, 8, 2, 6, 4	7. Nigeria	0,762	2, 6, 4
	7. Nigeria	4,98	5, 8, 2, 6, 4	3. Canada	1,398	5, 8, 2, 6, 4	3. Canada	0,711	2, 6, 4
	1. USA	4,88	8, 2, 6, 4	1. USA	1,227	5, 8, 2, 6, 4	1. USA	0,651	2, 6, 4
	5. India	4,69	6, 4	5. India	0,908	2, 6, 4	8. Peru	0,611	6, 4
	8. Peru	4,63	6, 4	8. Peru	0,779	4	5. India	0,441	6, 4
	2. UK	4,57	6, 4	2. UK	0,632	4	2. UK	0,313	4
	6. Malaysia	4,36		6. Malaysia	0,478	4	6. Malaysia	0,084	

	4. Germany	4,23		4. Germany	0,000		4. Germany	0,000	
Vertical Individualism	7. Nigeria	4,48	5, 6, 8, 3, 1, 2, 4	7. Nigeria	1,688	5, 6, 8, 3, 1, 2, 4	7. Nigeria	1,834	5, 6, 3, 8, 1, 2, 4
	5. India	3,74	3, 1, 2, 4	5. India	0,960	1, 2, 4	5. India	1,015	8, 1, 2, 4
	6. Malaysia	3,65	1, 2, 4	6. Malaysia	0,879	1, 2, 4	6. Malaysia	1,012	1, 2, 4
	8. Peru	3,51	4, 2	8. Peru	0,695	2, 4	3. Canada	0,729	4
	3. Canada	3,37	4	3. Canada	0,694	4	8. Peru	0,572	4
	1. USA	3,30	4	1. USA	0,562	4	1. USA	0,545	4
	2. UK	3,23	4	2. UK	0,395	4	2. UK	0,430	4
	4. Germany	2,89		4. Germany	0,000		4. Germany	0,000	
	RAW-Composite			**MG-CFA**			**ALIGNMENT**		
		Factor Mean	Groups With Significantly Smaller Factor Mean	-USA, Nigeria, and Peru; SDO08 ~1, SDO05 ~1, SDO04 ~1, SDO07 ~1	Factor Mean	Groups With Significantly Smaller Factor Mean		Factor Mean	Groups With Significantly Smaller Factor Mean
Social Dominance Ordination	7. Nigeria	3,75	6, 5, 2, 3, 8, 1, 4	6. Malaysia	0,934	2, 3, 4	7. Nigeria	1,297	5, 2, 8, 3, 1, 4
	6. Malaysia	3,31	2, 3, 8, 1, 4	5. India	0,768	2, 3, 4	6. Malaysia	1,117	2, 8, 3, 1, 4
	5. India	3,21	2, 3, 8, 1, 4	2. UK	0,373	4	5. India	0,925	2, 8, 3, 1, 4
	2. UK	2,70	4	3. Canada	0,288		2. UK	0,452	4
	3. Canada	2,64	4	4. Germany	0,000		8. Peru	0,403	4
	8. Peru	2,57	4				3. Canada	0,380	
	1. USA	2,51	4				1. USA	0,314	
	4. Germany	2,10					4. Germany	0,000	
	RAW-Composite			**MG-CFA**			**ALIGNMENT**		
		Factor Mean	Groups With Significantly Smaller Factor Mean	-India and Malaysia; RWA06 ~1, RWA01 ~1, RWA04 ~1, RWA02 ~1	Factor Mean	Groups With Significantly Smaller Factor Mean		Factor Mean	Groups With Significantly Smaller Factor Mean
Right Wing Authoritarianism	7. Nigeria	5,43	6, 5, 2, 8, 1, 3, 4	3. Canada	0,454	4, 3, 2, 8, 1	7. Nigeria	2,666	6, 5, 8, 2, 1, 3, 4
	6. Malaysia	4,71	2, 8, 1, 3, 4	4. Germany	0,000	1, 3, 4	6. Malaysia	1,876	8, 2, 1, 3, 4
	5. India	4,47	2, 8, 1, 3, 4	7. Nigeria	2,858	1, 3, 4	5. India	1,687	2, 1, 3, 4

	2. UK	3,44	1, 3, 4	8. Peru	1,138	4	8. Peru	0,919	1, 3, 4
	8. Peru	3,44	1, 3, 4	2. UK	1,024	4	2. UK	0,732	1, 3, 4
	1. USA	2,82	4	1. USA	0,646		1. USA	0,266	
	3. Canada	2,60					3. Canada	0,053	
	4. Germany	2,46					4. Germany	0,000	
	RAW-Composite			**MG-CFA**			**ALIGNMENT**		
		Factor Mean	Groups With Significantly Smaller Factor Mean	-India; VS9 ~1, VS7 ~1, VS6 ~1, VS8 ~1, VS3 ~1	Factor Mean	Groups With Significantly Smaller Factor Mean		Factor Mean	Groups With Significantly Smaller Factor Mean
Victim Sensitivity	7. Nigeria	4,15	5, 4, 1, 8, 6, 2	7. Nigeria	0,116	1, 6, 2	7. Nigeria	0,193	1, 6, 2
	5. India	3,95		4. Germany	0,000	2	4. Germany	0,000	
	4. Germany	3,93		8. Peru	-0,106		5. India	-0,003	
	3. Canada	3,92		3. Canada	-0,100		8. Peru	-0,053	
	1. USA	3,89		1. USA	-0,195		3. Canada	-0,085	
	8. Peru	3,87		6. Malaysia	-0,209		1. USA	-0,155	
	6. Malaysia	3,86		2. UK	-0,257		6. Malaysia	-0,214	
	2. UK	3,85					2. UK	-0,224	
	RAW-Composite			**MG-CFA**			**ALIGNMENT**		
		Factor Mean	Groups With Significantly Smaller Factor Mean	-Nigeria; OS6~1, OS1~1, OS3~1	Factor Mean	Groups With Significantly Smaller Factor Mean		Factor Mean	Groups With Significantly Smaller Factor Mean
Observer Sensitivity	8. Peru	4,40	6, 4, 1, 5, 3	8. Peru	0,387	6, 4, 5, 1, 3	7. Nigeria	0,330	6, 4, 1, 5, 3
	7. Nigeria	4,40	6, 4, 5, 1, 3	2. UK	0,223	3, 5, 1	8. Peru	0,313	6, 4, 1, 5, 3
	2. UK	4,25	1, 5, 3	6. Malaysia	0,049		2. UK	0,188	1, 5, 3
	6. Malaysia	4,14	3	4. Germany	0,000		6. Malaysia	0,021	
	4. Germany	4,14		5. India	-0,021		4. Germany	0,000	
	5. India	4,02		1. USA	-0,046		1. USA	-0,117	
	1. USA	4,01		3. Canada	-0,241		5. India	-0,152	
	3. Canada	3,90					3. Canada	-0,242	

Table V-14

Country values on additional variables

	Individualism-Collectivism	Power Distance	GNI $ in dollars	GNI world rank	Egalitarian value	Important in life: Religion
1. USA	91	40	58270	6	4,68	2,03
2. UK	89	35	40530	18	4,92	2,64
3. Canada	80	39	42870	16	4,80	2,25
4. Germany	67	35	43490	15	5,01	2,74
5. India	48	77	1820	140	4,45	1,41
6. Malaysia	26	100	9650	61	4,41	1,19
7. Nigeria	30	80	2080	136	4,79	1,13
8. Peru	16	64	5970	82	4,84	1,71

Comparison to established research

HC/VC/HI/VI. For these measures, we looked at the comparison to Hofstede's IND-COL and power distance scale. We found no relationship between these scales and IND-COL measure for both methods. The mean individualism and collectivism values for each country, regardless of horizontal of vertical dimension were uncorrelated to the values for each country reported by Hofstede et al., (2010). For power distance, the latent mean found with partial invariance was significantly positively correlated with vertical collectivism (see Table 15), and for the Alignment latent means with vertical individualism (see Table 16). Power distance was not correlated with the horizontal aspects of individualism and collectivism. We also found vertical individualism positively correlated with GNI world rank and importance of religion for both our latent mean values, meaning that the vertical components on the individual and collective scales are related as predicted to power distance and also other measure that capture societal hierarchy. In the comparison of methods, although both vertical collectivism and vertical individualism have elements that should be related to power distance, some literature suggests that vertical individualism in particular reflects Hofstede's conceptualization of power distance (Schimmack, Oishi, & Diener, 2005), pointing to the validity of the Alignment method.

SDO/RWA. As the literature would predict, we found SDO and RWA to be highly correlated at the country level. Looking at the GNI index, our findings are in line with the understanding of SDO for country income for alignment results but not for partial invariance results. We found no relationship to country level egalitarian values. Not explicitly predicted by the literature we reviewed, we also found a positive correlation to power distance and importance of religion for both latent means. We found a very similar pattern with RWA as SDO: positive correlation to power distance, importance of religion, and national income, but not in dollar amount in the partial invariant results. The prediction that SDO was significantly negatively correlated with GNI index was only true for the Alignment method, so although there are similarities, with specific predictions the Alignment method results were more in line with previous research.

VS/OS. According to the literature, we would expect no systematic cultural level differences and minimal differences overall for both VS and OS. This is what we found in our analysis; minimal differences between the countries, even more so with the alignment results. However, because so little research has been done in this area, a much greater number of countries and samples that are more diverse should be surveyed before we conclude that VS and OS show minimal variation on the country level.

Table 15

Correlation between country level latent mean found with MG-CFA and other country level variables

Partial Invariance	VS	OS	HC	VC	HI	VI	SDO	RWA	IND-COL	Power Distance	GNI $ in dollars	GNI world rank	Egali vaule	Religion
VS N = 7	1													
OS N = 7	-0,213	1												
HC N = 8	0,63	0,688	1											
VC N= 8	0,519	0,321	0,675	1										
HI N = 8	0,074	-0,384	-0,413	-0,183	1									
VI N = 8	0,457	-0,059	0,25	0,68	0,578	1								
SDO N = 8	0,338	-0,409	0,067	,740*	0,292	,825*	1							
RWA N = 6	0,497	0,712	0,509	,831*	0,411	,931**	0,811	1						
Individualism-Collectivism	-0,377	-0,493	-0,654	-0,663	0,148	-0,545	-0,415	-0,574	1					
Power Distance	0,152	0,164	0,282	,735*	0,003	0,698	,779*	,884*	-,826*	1				
GNI $ in dollars	-0,382	-0,418	-0,572	-,747*	0,014	-0,689	-0,607	-0,762	,901**	-,855**	1			
GNI world rank	0,606	0,246	0,452	,726*	0,181	,786*	0,685	,902*	-,724*	,743*	-,916**	1		
Egalitarian vaule	0,348	0,218	0,401	-0,239	-0,208	-0,441	-0,672	-0,324	0,34	-,752*	0,457	-0,445	1	
Important in life: Religion	0,256	0,079	0,176	0,648	0,325	,859**	,800*	,851*	-0,756	,931**	-0,806	,799*	-0,738	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Table 16

Correlation between country level latent mean found with Alignment and other country level variables

Alignment N = 8	VS	OS	HC	VC	HI	VI	SDO	RWA	IND-COL	Power Distance	GNI $ in dollars	GNI world rank	Egali vaule	Religion
VS	1													
OS	0,305	1												
HC	0,491	,920**	1											
VC	0,395	0,361	0,276	1										
HI	0,381	0,082	0,039	-0,251	1									
VI	0,534	0,283	0,18	0,664	0,498	1								
SDO	0,316	0,274	0,13	,799*	0,186	,925**	1							
RWA	0,438	0,478	0,355	,848**	0,139	,884**	,960**	1						
Individualism-Collectivism	-0,398	-0,538	-0,57	-0,436	0,005	-0,508	-0,587	-0,694	1					
Power Distance	0,202	0,246	0,148	0,671	-0,055	,711*	,872**	,865**	-,826*	1				
GNI $ in dollars	-0,452	-0,459	-0,486	-0,583	-0,075	-0,666	-,759*	-,841**	,901**	-,855**	1			
GNI world rank	0,647	0,351	0,377	0,617	0,267	,762*	,771*	,854**	-,724*	,743*	-,916**	1		
Egalitarian vaule	0,188	0,35	0,507	-0,344	0,029	-0,465	-0,649	-0,516	0,34	-,752*	0,457	-0,445	1	
Important in life: Religion	0,345	0,223	0,101	0,568	0,29	,847**	,888**	,869**	-0,756	,931**	-0,806	,799*	-0,738	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

VS = Victim Sensitivity OS = Observer Sensitivity HC = Horizontal Collectivism VC = Vertical Collectivism HI = Horizontal Individualism VI = Vertical Individualism SDO = Social Dominance Ordination RWA = Right Wing Authoritarianism

Discussion

This study looked at data collected in eight different cultures to compare the utility of two different methods to test measurement invariance. Literature in measurement equivalence is clear in that the lack of scalar MI can lead to erroneous conclusions in comparisons of the composite means, and in the case of this data set, the mean difference from the raw scores was inflated. Lacking scalar invariance, we looked at the latent means derived from two methods. Past research has been mixed on the utility of partial invariance for comparing latent means; adding to this debate, we found differing results between latent mean comparisons using MG-CFA and Alignment methods. In six out of eight scales, the MG-CFA method found mean differences where the Alignment method did not. The remaining two found the same number of significant differences. In Marsh et al. (2016), the difference of significance between groups comparing latent means estimated on the basis of MG-CFA (with partial invariance) compared to on the basis of Alignment also shifted, but the direction depended on the measure. Simulation studies point to evidence that the relative bias of partial MG-CFA compared to approximate methods is sensitive to the number and size of differences in intercepts as well as the sample size (Marsh et al., 2016; van de Schoot et al., 2013).

The results we found here are in line with the idea that partial invariance is problematic, in part because the stepwise procedure is post-hoc and there is no guarantee that the process based on freeing parameters to achieve acceptable fit will identify the true model and unbiased means (Marsh et al., 2016). More research in simulation studies is needed to confirm that the Alignment method gives less biased means compared to MG-CFA. In addition, these limitations could lead to the lack of replication for real data: no studies that we know have looked at replication of results with real data using these methods. If the scales are in fact comparable as suggested by our results, we should find invariance with new data collected in similar samples. According to the partial invariance limitations, finding the same model with the same modification as the partial invariance solution is dubious. However, the

Alignment method contains no such limitations. In addition to the question around the validity of partial invariance model, for half the scales, a full comparison was not possible with MG-CFA because we had to drop a country or more; this was not a problem with the Alignment method.

The demonstration fostered the understanding of differences on a theoretical level by looking at the results in the context of past research on cross-cultural difference at a country level. This was a limited endeavor because research in the area of country level differences for our variables is incomplete. The results in this analysis for the two methods were quite similar; however, where differences were found, our results for the Alignment method fit the general patterns found in past research better than those with partial invariance. Somewhat surprisingly, it was difficult to decipher a pattern for most of the cross-culturally studied variables, namely HC/VC/HI/VI, and additionally, they showed no relationship to Hofstede's IND-COL for either method. However, it is not completely unforeseen because literature has long called for a reexamination of individualism and collectivism constructs as a whole, particularly Hofstede's approach (McSweeney, 2002). Specifically, the variables we measured are thought of in the literature as an individual level variable and not researched widely at country level. Hofstede's power distance, on the other hand, was correlated with dimensions of the vertical scales as predicted; however, which dimension was significantly correlated differed with the method of analysis. The Alignment results showed a significant correlation with vertical individualism which is a better reflection on Hofstede's conceptualization of the construct. Although the results for SDO and RWA were similar for the two methods, the predicted relationship to national income measure by GNI was found only in the Alignment method results. Finally, in VS and OS, we did not predict country level mean difference scores, and we in fact found less significant mean difference score using the Alignment method. Although these results show a stronger validation the Alignment method

analysis, more research is needed before we can understand these variables at the country level.

Limitations

A limitation to the method in this study is the sample size and sample makeup. Firstly, in methods validation, larger sample sizes are sometimes recommended for MI testing (Meade, 2005; Meade & Lautenschlager, 2004). Secondly, to be able to make a definitive statement about the country level means, it is necessary to have a representative sample, whereas our sample consists of mostly young adults and students. However, the samples we obtained are representative of cross-cultural research. Often, when involving many different research groups in a project, there is a disparity in what collaborators deliver, and less than ideal sample sizes is the norm. Although a majority of undergraduates in the samples is a limitation in our conclusions about country level means, having sample groups that are similar in age and education and only differ in nationality leads to more conclusive findings with regard to the equivalence of the measurement scales used. Furthermore, both of these points are strengths as well as weaknesses because they represent common research practice, and lie at the heart of our question which is the utility of method for use in common research. Future research should both replicate these findings and compare them to further alternative methods.

An additional caveat to all invariance testing is that it can only statistically insure statistical equivalence. Even if we can establish strict MI confirming that there is no bias in factor loadings and item intercepts across countries, we cannot guarantee conceptual equivalence or, in other words, lack of construct bias. The whole construct could differ in meaning across cultures. MI does not necessarily mean lack of all types of bias. For example, bias could influence all the indicators to the same degree, therefore having a biased estimation of the latent mean. Future research should investigate how to theoretically identify measurement bias that is not statistically identifiable.

In our contextualization of the results, although we looked at the correlation of the country level variable to provide context to our findings, the sample size here of 8 countries is not large enough to make any definitive conclusions. Much larger research projects should be implemented in this regard. However, this was a meaningful analysis because these results gave theoretical support to our methodological exercise. Furthermore, although not definitely conclusive, they support the assumption that Alignments results are more trustworthy.

Conclusion

Establishing measurement invariance is necessary for the comparison of homogeneous groups; however, it is often overlooked. Research has highlighted the limitations of the MG-CFA for large scale studies and opened the door to exploring new methods for testing MI. This paper sheds light on the fact that this is also important for medium scale studies. Past solutions of limiting the countries in analysis of partial invariance are insufficient. This article demonstrates the strengths of Alignment method compared to MG-CFA, in that it is more flexible and therefore more likely to find invariance without the multiple corrections required with partial invariance, and still has less biased latent means. Furthermore, we found that across measures, the MG-CFA shows consistently more significant mean differences than the Alignment method. This difference in itself supports the argument that partial invariance has many significant limitations. Future research should replicate these findings and further explore what this means on a conceptual level.

References

Altemeyer, B. (1981). *Right-wing authoritarianism*. Winnipeg, Canada: University of Manitoba Press

Altemeyer, B., & Hunsberger, B. (1992). Authoritarianism, religious fundamentalism, quest, and prejudice. *The international journal for the psychology of religion*, *2*, 113-133.

Asparouhov, T., & Muthén, B. O. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–14.

Boer, D., Hanke, K., & He, J. (2018). On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. *Journal of Cross-Cultural Psychology*, *49*, 713-734.

Byrne, B. M., & van de Vijver, F. J. (2017). The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema*, 29.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*, 464–504.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*, 1005-1018.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescent response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*, 187-212.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.

Cohen, J. (1994). The earth is round (p<. 05). In What if there were no significance tests? (pp. 69-82). Routledge.

Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology*, *43*, 558-575.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*, 55–75.

Duckitt, J., & Sibley, C. G. (2007). Right wing authoritarianism, social dominance orientation and the dimensions of generalized prejudice. *European Journal of Personality: Published for the European Association of Personality Psychology*, *21*, 113-130.

Hofstede, G. (1980a). *Culture's Consequences: International Differences in Work-Related Values.* Beverly Hills, CA: Sage.

Hofstede, G., Hofstede, G. J. & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind* (Rev. 3rd ed.). New York: McGraw-Hill.

Heath, A., Martin, J., & Spreckelsen, T. (2009). Cross-national Comparability of Survey Attitude Measures. *International Journal of Public Opinion Research*, *21*, 293–315.

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., Foels, R., & Stewart, A. L. (2015). The Nature of Social Dominance Orientation: Theorizing and Measuring Preferences for Intergroup Inequality Using the New SDO7 Scale. *Journal of Personality and Social Psychology*, *109*, 1003-1028.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014. World Values Survey: Round Six - Country-Pooled Datafile Version: http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp. Madrid: JD Systems Institute.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin & B. Puranen et al. (eds.). 2014. World Values Survey: Round Five - Country-Pooled Datafile Version: www.worldvaluessurvey.org/WVSDocumentationWV5.jsp. Madrid: JD Systems Institute.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). semTools: Useful tools for structural equation modeling. R package version 0.5-1. Retrieved from https://CRAN.R-project.org/package=semTools

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*, 524-544.

Lomazzi, V. (2018). Using Alignment Optimization to Test the Measurement Invariance of Gender Role Attitudes in 59 Countries. *Methods, data, analyses: a journal for quantitative methods and survey methodology (mda), 12,* 77-103.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504.

Magraw-Mickelson, Z., & Gollwitzer, M. (2018). Relational and Group Collective Self Responses to Observed Victimization Across Cultures. *Social Justice Research, 31,* 113-132.

Manganelli Rattazzi, A. M., Bobbio, A., & Canova, L. (2007). A short version of the Right-Wing Authoritarianism (RWA) Scale. *Personality and Individual Differences*, *43*, 1223-1234.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, *98*, 224.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2016). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*, 524-545.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling*, *11*, 320-341.

McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith-a failure of analysis. *Human relations*, *55*, 89-118.

Meade, A. W. (2005, April). Sample size and tests of measurement invariance. In Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling*, *11*, 60-72.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.

Muthén, B., & Asparouhov, T. (2013a). BSEM Measurement Invariance Analysis. Mplus Web Notes: No. 17, January 11. (Vol. 17, p. 313). Retrieved February 2, 2017, from https://www.statmodel.com/examples/webnotes/webnote17.pdf

Muthén, B., & Asparouhov, T. (2013b). New Methods for the Study of Measurement Invariance with Many Groups. Technical report. Available online at: http:// statmodel2.com/download/PolAn.pdf.

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology, 5,* 978.

Muthén, L. K., & Muthén, B. O. (1998–2018). *Mplus user's guide (version 8.1).* Los Angeles, CA: Muthén & Muthén.

Perry, R., Sibley, C. G., & Duckitt, J. (2013). Dangerous and competitive worldviews: A meta-analysis of their associations with social dominance orientation and right-wing authoritarianism. *Journal of Research in Personality, 47,* 116-127.

Pratto, F., Liu, J. H., Levin, S., Sidanius, J., Shih, M., Bachrach, H., & Hegarty, P. (2000). Social dominance orientation and the legitimization of inequality across cultures. *Journal of Cross-Cultural Psychology, 31,* 369-409.

Pratto, F., Sidanius, J., Stallworth, L., & Malle, B. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67,* 741–763.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552-566.

Rensvold, R. B., & Cheung, G.W. (2001). Testing for metric invariance using structural equation models: Solving the standardization problem. In C. A. Schriesheim & L. L. Neider (Eds.), Research in management: Vol. 1. Equivalence in measurement (pp. 21-50). Greenwich, CT: Information Age.

Rosseel, Y (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48,* 1-36. URL http://www.jstatsoft.org/v48/i02/

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74,* 31-57.

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity. *European Journal of Psychological Assessment*, *21*, 202-211.

Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity: Assessment and location in the personality space. *European Journal of Psychological Assessment*, *21*, 202–211.

Sibley, C. G., & Duckitt, J. (2008). Personality and prejudice: A meta-analysis and theoretical review. *Personality and Social Psychology Review*, *12*, 248-279.

Sörbom (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229-239.

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78-90.

Steinmetz, H. (2013). Analyzing observed composite differences across groups. *Methodology. 9*, 1-12.

Schwartz, S. H. (1994). Beyond individualism-collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, S. C. Choi, & G. Yoon (Eds.), Individualism and collectivism: Theory, method and applications (pp. 85–119). Thousand Oaks, CA: Sage.

Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative Sociology, 5*, 136–182.

Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of personality*, 69, 907-924.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology, 74*, 118.

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthen, B. (2013). Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology, 4,* 770.

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *54*, 119–135.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4–70.

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational research methods*, *5*, 139-158.

World Bank (2008). World Bank, International Comparison Program database. Retrieved 1 December 2018 from https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD

Wu, M. S., Schmitt, M., Zhou, C., Nartova-Bochaver, S., Astanina, N., Khachatryan, N., & Han, B. (2014). Examining self-advantage in the suffering of others: Cross-cultural differences in beneficiary and observer justice sensitivity among Chinese, Germans, and Russians. *Social Justice Research*, *27,* 231-242.

Jang, S., Kim, E. S., Cao, C., Allen, T. D., Cooper, C. L., Lapierre, L. M., ... & Abarca, N. (2017). Measurement invariance of the satisfaction with life scale across 26 countries. *Journal of Cross-Cultural Psychology*, *48,* 560-576.

Flake, J. K., & McCoach, D. B. (2018). An Investigation of the Alignment Method With Polytomous Indicators Under Conditions of Partial Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 56-70.

Fischer, R., Hanke, K., & Sibley, C. G. (2012). Cultural and institutional determinants of social dominance orientation: A cross-cultural meta-analysis of 27 societies. *Political Psychology, 33,* 437-467.

Schimmack, U., Oishi, S., & Diener, E. (2005). Individualism: A valid and important dimension of cultural differences between nations. *Personality and Social Psychology Review, 9,* 17-31.

Appendix A

Syntax for Mplus analysis victim sensitivity measure

```
Title: Alignment 1
Data: File IS "C:\Users \Documents\Mplus\all_Mplus.csv";
Variable: NAMES ARE ID sample vs01 vs02 vs03 vs04 vs05 vs06 vs07 vs08 vs09
    vs10
USEVARIABLES ARE vs01 vs02 vs03 vs04 vs05 vs06 vs07 vs08 vs09 vs10;
MISSING = ALL (9999);
CLASSES = c(8);
KNOWNCLASS = c(sample);
ANALYSIS: TYPE = MIXTURE;
ESTIMATOR = ML;
ALIGNMENT = fixed(4);
ASTARTS = 100
MODEL: %OVERALL%
F1 by vs01 vs02 vs03 vs04 vs05 vs06 vs07 vs08 vs09 vs10;
OUTPUT: TECH1 TECH8 ALIGN;
```

VI. Chapter 6: Conclusion

The proceeding chapters outlined four studies that looked at cross-cultural social psychology, theory and methods. Part one focused on understanding the theory of individualism and collectivism, which is the most studied theory in the field. Part two looked at two areas for improvement in cross-cultural research methods. The latter two methodological studies draw directly from the challenges and limitations brought up in the studies looking at theory. Specifically, with the study in Chapter two, we were not able to compare some measures across countries because of the lack of measurement invariance. Chapter 5 proposes a new method of measurement invariance that shows promise for expanding the possibility of comparison using approximate rather than exact equivalence. Chapter 4 dealt with a problem that is often hidden in the literature, that is the quality of data and if this varies by data collection form and culture. Chapter 3 presented methodological problems that were not addressed in the other chapters but that should be a focus of future work, namely measurement and manipulation of self-construal.

Summary

Part 1

Part 1 includes two chapters that take different approaches to the theoretical clarification in what it means to include others in one's self-representation and how that is reflected in concern for the injustice of others. For the explanation of the limitation of the theoretical understanding of individualism and collectivism, I took the approach first proposed by Brewer and Gardner, (1996) and Kashima et al., (1995), that highlighted the importance of differentiating the relational self. While the overall concept of individualism and collectivism is the degree to which people identify with the group, the relation self clarifies the group as the connection to those with which one has a personal relationship.

Brewer and Chen (2007) are explicit in their explanation of the contrast between relational self and the group collective self, which is the importance of people and groups with whom one shares a social categorization.

Results of Chapter 2 show that over in the three countries sampled, there was a universal hierarchy for the importance of the relational self over the group collective self. That is, when orthogonally manipulated, participants across nationalities reported more vicarious anger and outrage as well as vicarious vengeful intentions when participants were told they knew the victim personally compared to when they did not. On the other hand, we did not find a main effect of group collective connection; participants did not show more anger and outrage or vengeful intentions when the victim was a member of their categorical group, specifically a colleague from the same department. In this study, we also did additional analysis to investigate the moderation of individual difference in measured self-representation. We found that in the German sample, there was a marginally significant group collective connection by individual self-representation interaction effect. Meaning, German participants high in individual self-representations reported more moral outrage when the victim had a collective group connection compared to when they did not.

It is relevant to note here that we specifically measured self-representations with Brewer and Chen's (2007) scale of individual, relational, and collective self-representation; we did not, however, find measurement invariance for this scale across countries, limiting our cross-cultural comparison. Furthermore, model fit was not very good, especially for the collective self-representation. This led to the decision not to use this scale in the next study, focusing on the more established scales of independent and interdependent self-construal. Additionally, the next study shifted to highlight the differentiating attributes of the group collective self namely shared group categorization. Therefore, the terminology changes to refer to the categorical self.

The study reported in Chapter 3 was intended to follow up the results found in Chapter 2 in the German sample with regards to the interaction between individual difference and group type. Conducting a lab study in China, we both measured self-construal and experimentally manipulated it using a priming technique. Priming can be useful in establishing causality of cultural variables; however, it was also limiting in this situation because we could not clearly say if the priming manipulation in fact primed self-construal. For this study, we considered previous studies that primed self-construal and the corresponding manipulation check. The measures were selected based on this thorough literature review; however, that does not guarantee a successful prime. The discrepancies between our prime manipulation and manipulation check result could be due to three things that are unrelated to the hypothesis of the study: 1. There is a publication bias in the literature that does not reflect the true nature of priming self-construal, in that, studies which were unsuccessful are not published, skewing the perception that priming self-construal is possible and 'works' as expected. 2. Priming procedures are fleeting in that by the time participants arrived at the priming manipulation check, the effect already wore off. 3. Research in priming self-construal using pronoun circling task and "Who am I?" manipulation checks have primarily been done in Western cultures or in English – Oyserman and Lee's (2008) meta-analysis reported only one study conducted in Chinese – therefore lack of expected relationship could be because of cultural differences in the method.

Chapter 3 further builds on Chapter 2 by exploring different conceptualizations of relational and categorical self. Focusing on the differentiation of defining groups by a relationship or category, we utilized the minimal group paradigm to establish two types of groups. First, drawing on the minimal group literature, participants were assigned a categorical group, "group A"; therefore, the categorical in-group members were define as other people in group A. Second, we tried a novel approach to connect strangers through a personal relationship. Participants were given the opportunity to talk to each other and discuss

personal connections; therefore, the relational in-group members were defined as other people in the same experimental session.

With this design, we did not find evidence for our many hypotheses with the priming manipulation. However, we did find some evidence looking at the measured self-construal in that those low in interdependent self-construal showed more emotional outrage when a member from the participants categorical group was treated unfairly compared to those high in interdependent self construal. Although these finding should be treated cautiously, they are consistent with the first study in that the categorical group is relatively more important to people for whom the individual self is relatively important, or in other words, those for whom the relational self is relatively unimportant. Much more research is needed to clarify the relationship between these scales or develop a new scale that more clearly reflects the conceptualization of differences in self-construal.

With the novel approach to relational minimal groups, it is not clear if the manipulation worked. The marginal results in relation to victim sensitivity point to some interaction between the relational minimal group and priming type, but overall the results are ambiguous. Future research that uses this method should include more comprehensive manipulation check items, for example, follow up questions in which the participant are asked what was discussed and what kind of connection they have.

The context of these two chapters was the observation to the injustice of others. I utilized this context because considering that how injustice befalling others affects the individual's emotional reaction is a concise way to determine who is relevant to the self. Due to this context, both studies also include other justice related variables. Specifically, predicating victim sensitivity, which is perceived injustice at one's own disadvantage (Gollwitzer, Schmitt, Schalke, Maes, & Baer, 2005), would moderator our findings. In this regard, these two studies found somewhat conflicting results. In Chapter 2, we found that those high in victim sensitivity were more outraged when there was a relational connection

compared to when there was no relational connection. In Chapter 3, we found a marginally significant interaction between victim sensitivity, group membership type, and relational self-construal prime, finding people high in victim sensitivity were relatively less emotional outraged when primed with relational self-construal and a member of their relational group was the victim of injustice. This discrepancy between the two studies may be due to the fact that in the second study, we measured relative emotional outrage of another person's victimization compared to the participant's victimization. Therefore, we can hypothesize that in situations of hypothetical or minimal groups, people high in victim sensitivity may have heightened outrage for relevant groups; however, not when in comparison to their own injustice. Future research should further explore when people high in victim sensitivity are also sensitive to the injustice of group members in experimental settings, including a comparison to their own disadvantage.

Part 2

Part 2 investigated cross-cultural psychology from a different angle. As outlined in the introduction and in the context of the studies in part 1, there are many methodological challenges we need to tackle before we can confidently compare cultures. The study in Chapter 4 emerged out of frustration with the collection of cross-cultural data and encountering issues with data quality. Observations of data collection lead me to theorize the possibility of cross-cultural differences in this regard. The goal of this study was to take advantage of research in the assessment of data quality (Curran, 2016; Meade & Craig, 2012) to reexamine the role on technology in data collection across various cultural groups. In light of the role of individual difference in careless responding (Bowling et al., 2016; Ward et al., 2017), the within-participant design made it possible to attribute differences in careless responding to the different modes, be that paper/pencil, computer/web-based, or smartphone, in which participants completed the survey. In this regard, we found differences between modes for some cultural groups: the sample of Chinese adults and German students, but not in

the sample of Chinese students. This brings attention to the importance of looking at culture not only across national borders but also subgroups with a national culture. Age groups within countries were particularly relevant for this research question because it considered participants' interaction with technology, an area in which there are understandable generational differences. Although these results were insightful, this study was not designed to definitively answer the question about the degree of careless responding between cultural groups.

Chapter 5 took a different approach to the methodical issue in cross-cultural research, also directly out of experience, which are the challenges with measurement invariance. Measurement invariance allows us to test if the measurement scales we use in heterogonous groups are comparable. In the study in Chapter 2, we tested measurement invariance for all measures; for our main dependent variable of emotional outrage, we found our measure was comparable and subsequently compared the degree of moral outrage across countries. This comparison was not possible for the measurement of self-representations. This was the limitations for our three group sample using MG-CFA. With this small number of groups, this method of invariance testing was possible and desired. However, as the number of groups increases so does the difficulty of establishing full invariance even if the measurement scales are practically comparable (Marsh et al., 2017; Rutkowski & Svetina 2014). Rather than changing the analysis plan, as in the Chapter 2 study, or dropping countries from analysis as necessary in Lomazzi (2018), in which only 27 out of 59 countries were comparable, it is of utmost importance to understand new methods. This was the goal in Chapter 5: to compare Alignment method to partial invariance using MG-CFA. Results show that Alignment method is a promising alternative to partial invariance.

Reflecting on these two distinct methodological considerations together, we can see that perhaps they are complimentary. Preliminary results (Magraw-Mickelson, 2018) showed that when we looked at the longitudinal invariance for the three waves of data collection, we

found invariance across waves for the Chinese adult sample (other samples were not large enough for this analysis) only when we exclude the 10% most careless responders. This is a promising avenue for the verification for careless responding indices; however, this was not included in the final manuscript in Chapter 4 because the 10% exclusion cut off was arbitrarily chosen. On the other hand, we did not check for careless responding in the measurement invariance study of Chapter 5. If there are significant differences between countries in careless responding, that may account for any between cultural differences. We did not measure careless responding because the data for this study was not suitable for a multi-dimensional index. Future research should clarify the utility if only select indices are identifiability as in Chapter 5 data set. However, the results are unlikely due to cultural difference in careless responding alone because of our overall findings of comparability to existing constructs in the literature. Nonetheless, the possibility of some noise due to careless responding cannot be discounted. Future cross-cultural research should both verify the quality of data with careless responding indices and the comparability with testing for measurement invariance.

General Contributions

In addition to taking on some of the big issues of the field, this body of work further contributed to the field of CCSP by not depending on past assumptions and being open to the unknown. That is, this research does not just assume cross-cultural differences but also uncovers cross-cultural similarity, looking at within-culture variations as well as between-culture, and finally, asks fundamental methodical questions for any cross-cultural study. In Chapter 2, although the background of the framework in self-construal is an East-West difference, it was important to know that when considering the overall hierarchy of the self, there is actually cross-cultural similarity. This can contribute to the literature on the self and also how to approach differences, which perhaps occur not on the cultural level but the individual level within countries. Building on that, Chapter 3 looked at differences in self-

construal within China. The field acknowledges that there is a lot of national heterogeneity within cultures (i.e. McSweeney, 2002; Voronov & Singer 2002); however, this can often be minimized when the goal of research is a cross-national comparison. Furthermore, when studies look at within-country psychological differences, this is most often within Western countries, going back to the issue of over reliance on WEIRD samples (Arnett, 2008; Cross et al., 2011; Henrich, Heine, & Norenzayan, 2010). The study in Chapter 3 fills this gap by not only looking at differences within a non-Western culture, but also both measured and manipulated self-construal. Part two contributes to the literature by reexamining fundamental methodological assumptions. From the review in Chapter 4, it is evident that much of the research in the field has presumed modes of data collection that are essentially equal, and very few have looked at data quality and culture in conjunction. The results from this study upend this assumption and fill the gaps in the literature. Finally, Chapter 5 addresses the vital topic of measurement equivalence, shedding light on the fact that depending on partial invariance might be a flawed strategy and thus, introduces an alternative.

General Limitation and Future Direction

Although these are important contributions, there are also some limitations as these studies do not and cannot address all issues in full. I believe what is missing from these studies is what needed to further strengthen researchers' approach to the field of cross-cultural psychology as a whole, beyond addressing the major topics introduced in Chapter 1, that is: 1) replications, 2) alternative methods, and 3) better scientific practices. I will first discuss how these apply to cross-cultural psychology as a whole and then individually for the studies.

Central to strengthening the field is to understand what is so central to the replicability crisis debate in the first place, that is, replication. While social psychology as a whole focuses on the replication of fundamental findings, very little replication work has looked at cross-cultural research. Part of the replication process for CCSP should be an expansion of

populations of interest, moving away from the over reliance on one country to represent a whole region and the focus on East-West comparisons.

Also, it is problematic that, like many social psychology areas, cross-cultural social psychology is overly dependent on surveys to measure psychological constructs. It has long been recognized that it is important to verify results with different methods (Gelfand, 2012; Cohen, 2007). With CCSP, this is even more vital because each method has unique comparability drawbacks. As Gelfand (2012) says, "In cross-cultural research, all methods have additional cultural baggage" (p. 191). Therefore, findings in CCSP should be confirmed with more than one method, whenever possible.

In the wake of the replicability crisis, in addition to the call for more attention to context and boundary conditions, there is also a heightened focus on better research practices; for example, the push for preregistration. Preregistration requires researchers to explain their hypotheses ahead of time, including methods, and analysis plan with the goal of increasing the credibility of research. This is beneficial to research because it prioritizes theory and method, distinguishes exploratory and confirmatory analysis, and reduces publication bias (van't Veer, & Giner-Sorolla, 2016). These goals are also of utmost importance to cross-cultural psychology, in the sense that, as outlined in the introduction, the strength of the most commonly used theories such as Hofstede's dimensions is an ongoing issue. Furthermore, I believe that cross-cultural psychology has an acute problem with publication bias, in that when results do not fit the IND-COL foundational hypothesis or when methods do not show differences, studies are not published. Practices like preregistration or data banks for null findings will help the field make progress and avoid pitfalls such as data 'fishing' and the 'file drawer' problem.

These general limitations, namely lack of replications, diverse methods and preregistrations also apply to the present body of work. I will reflect on the present studies in the context of these limitations and talk about the best ways forward.

Replication

I believe replication is the most important aspect that can strengthen this body of work. For Part 1, although Chapter 2 is a follow up of Chapter 3, in that it addresses the meaning of collectivism and the within-country difference in self-construal proposed by Chapter 2, they are different studies and the core findings of both chapters still need to be replicated. First, with Chapter 2, we found similarity across cultures; however, this study, which includes three countries, only had one from a non-Western sample. As Reid (2002) reminds us, one group or sub-group is not representative of a whole region or country. Just because the Japanese are similar to Americans and Germans in the hierarchy of the self, we cannot assume that in other non-Western countries we would find the same thing. Cross-cultural research studies should always strive to include more countries.

This is also seen in Chapter 4, which looked at only German and Chinese cultures. By adding age groups in the Chinese sample, the results demonstrated that a particular demographic group within a culture alone does not establish cultural norms. Future research on data quality should also include different age groups in Germany as well as other regions. Although Chapter 3 compares within-country differences in self-construal, the results should be replicated in another culture; data is currently being collected in Germany to this end. In addition, the findings would be strengthened with a conceptual replication, specifically using a different context of 'unfair' treatment; this is also taking place in an ongoing online study that looks at a different context and conceptualization of relational and categorical self.

In part 2, Chapter 4 shows the importance of replication in that is reexamined the role of data collection mode for data quality. However, that opens the questions about of the rate of change, so that we must ask our self questions such as: How quickly does humans relationship to technology change in a way that will influence data quality? At what point will these results also be obsolete?

The results of Chapter 5 will be robust with the addition of an exact replication. The measurement invariance methods have not been compared with real data in the past, so replicating the findings is of utmost importance. Therefore, the next planned study will collect data in the same countries, with similar convenience samples, but with at least 200 participants per sample.

Diverse methods

The above studies could also use more dimensionality in the methods. The dependency on surveys was supplemented by the methods section's examination of survey comparability; however, there is still room for improvement. Chapter 3 was interesting in that it was a lab study that experimentally manipulated variables rather than only measurement with surveys; however, in the end, the priming manipulations did not work as well as was hoped. It is unclear if the priming, the priming manipulation check, or neither, were actually successful. Nonetheless, we learned that even though these are the most common methods for this type of research, they are also severely limited and new methods need to be explored.

The conclusions of Chapter 4 had two limitations, both of which would benefit from an exploration of new methods. First, we found differences between modes but could not use psychological variables to explain those differences. Future studies should find different ways to measure personality variables that do not require responses that may have been written carelessly. Second, we could not describe the differences in the degree of careless responding definitively between countries and could not recommend a cutoff point. New methods should be formulated to tackle this issue.

In this area of expanding methods, the one topic from the big questions described in the introduction that this set of studies was not able to touch on was the understanding and betterment of existing measures of self-construal. Future research should investigate questions related to the dimensionality of constructs, the relationship between existing measure, and if there is a better way to measure self-construal.

Preregistration

Finally, in the interest of strengthening credibility with better research practices, only one of the studies, Chapter 4, was preregistered. Preregistration was especially important for that chapter because it was the most confirmatory of all the studies. Nonetheless, the other studies could have also been bolstered through these means. Chapters 4 and 5 were exploratory in nature, however, preregistration can also document that this is the case. It is most important that preregistration takes place in the next steps of replication.

Overall, this collection of studies was progressive in confronting fundamental issues in the field of cross-cultural psychology, both in terms of theoretical framework and confronting methodological limitations. However, there are still improvements to be made, as these are only the first steps in what is a long journey.

Conclusion

In conclusion, through a wide variety of studies, this body of work drew together an extensive sectional of important topics in cross-cultural social psychology. The contributions of this work are many: 1) Finding similarity in the hierarchy in self-representations where cross-cultural theories would assume differences. 2) Using a lab study whereas a great deal of research is done by survey and establishing the use of relational minimal groups. 3) Investigating culture and careless responding together and finding difference between modes that varied between national culture and age group. 4) Finally, comparing old and new methods to establish equivalence which is necessary for any group comparisons. The steps taken here are important and, although there is more to be done, each step is significant on the road to a better understanding of cross-cultural social psychology.

References

Arnett, J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist, 63,* 602-614.

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology, 111,* 218-229.

Brewer, M. B., & Chen, Y. R. (2007). Where (who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychological review, 114*, 133-151.

Brewer, M. B., & Gardner, W. (1996). Who is this "We"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology, 71,* 83–93.

Cross, S. E., Hardin, E. E., & Gercek-Swing, B. (2011). The what, how, why, and where of self-construal. *Personality and Social Psychology Review, 15*, 142-179.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19.

Gelfand, M. J. (2012). The Trials and Tribulations of Cross-Cultural Research. Ryan, Ann Marie (Ed); Leong, Frederick T. L. (Ed); Oswald, Frederick L. (Ed), (2012). Conducting multinational research: Applying organizational psychology in the workplace. *American Psychological Association*. (ch. 7, p. 179-199)

Gollwitzer, M., Schmitt, M., Schalke, R., Maes, J., & Baer, A. (2005). Asymmetrical Effects of Justice Sensitivity Perspectives on Prosocial and Antisocial Behavior. *Social Justice Research, 18,* 183–201.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences, 33,* 61-83.

Kashima, Y., Yamaguchi, S., Kim, U., Choi, S., Gelfand, M. J., & Yuki, M. (1995). Culture, gender, and self: A perspective from individualism-collectivism research. *Journal of Personality and Social Psychology, 69,* 925-937.

Lomazzi, V. (2018). Using Alignment Optimization to Test the Measurement Invariance of Gender Role Attitudes in 59 Countries. *Methods, data, analyses: a journal for quantitative methods and survey methodology (mda), 12,* 77-103.

Magraw-Mickelson, Z. (2018, July). Survey Mode and Data Quality: A Look at Careless Responding Across Three Modes and Two Cultures. International Association of Cross-Cultural Psychology Congress, Guelph, Canada.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2017). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23,* 524-545.

McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith-a failure of analysis. *Human relations, 55*, 89-118.

Meade, A.W. & Craig, S.B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17, 437- 455.

Oyserman, D., & Lee, S. W. S. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological Bulletin, 134*, 311-342.

Reid, P. T. (2002). Multicultural psychology: Bringing together gender and ethnicity. Cultural *Diversity and Ethnic Minority Psychology, 8,* 103-114.

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*, 31-57.

van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2-12.

Voronov, M., & Singer, J. A. (2002). The myth of individualism-collectivism: A critical review. *Journal of Social Psychology*, *142*, 461-480.

Ward, M. K., Meade, A. W., Allred, C. M., Pappalardo, G., & Stoughton, J. W. (2017). Careless response and attrition as sources of bias in online survey assessments of personality traits and performance. *Computers in Human Behavior*, *76*, 417-430.

Acknowledgement

First and foremost my deepest gratitude to Prof. Dr. Mario Gollwitzer, his guidance and support made this possible. I feel very fortunate to be able to work under his supervision; I have learned so much on how to think about psychology and how to be a better researcher, all of which helped strengthen this dissertation in every way.

I wish to dedicate this to my parents who have always believed in me and my abilities to accomplish great things.

Heartfelt thanks to Megan Gilmore and Raudhah Fatin Hirschmann for all the help over the years, I could not have done it without them. Also, for my other helpers who gave assistance and support along the way: Shankar, Julie, Annie, Naomi, Nicole, colleagues: Johannes, Gunnar, Merle, Mathias, and the rest of my Marburg crew: Corey, Shuai, Vanessa. Furthermore, my sincerest gratitude to my collaborators on the projects presented in this dissertation.

Finally, I appreciate all of the emotional support from friends and family around the world. Thank you for the encouragement and inspiration.

Relevant Publications

Magraw-Mickelson, Z., & Gollwitzer, M. (2018). Relational and Group Collective Self Responses to Observed Victimization Across Cultures. *Social Justice Research, 31*, 113-132.

Magraw-Mickelson, Z., Wang, H., & Gollwitzer, M. (2019). Survey Mode and Data Quality: A Careless Responding Across Three Modes in Cross-Cultural Contexts. *Manuscript submitted for publication.*