CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Universal Dimensions of Individuals' Perception

Revisiting the Operationalization of Warmth and Competence with a Mixed-Method Approach

Halkias, Georgios; Diamantopoulos, Adamantios

Document Version
Final published version

Published in:
International Journal of Research in Marketing

DOI:
10.1016/j.ijresmar.2020.02.004

Publication date:
2020

License
CC BY

Citation for published version (APA):
Halkias, G., & Diamantopoulos, A. (2020). Universal Dimensions of Individuals' Perception: Revisiting the Operationalization of Warmth and Competence with a Mixed-Method Approach. *International Journal of Research in Marketing*, *37*(4), 714-736. https://doi.org/10.1016/j.ijresmar.2020.02.004

Link to publication in CBS Research Portal

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 10. Jul. 2025

International Journal of Research in Marketing 37 (2020) 714–736

Contents lists available at ScienceDirect

IJRM

International Journal of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

Universal dimensions of individuals' perception: Revisiting the operationalization of warmth and competence with a mixed-method approach

Georgios Halkias *, Adamantios Diamantopoulos

University of Vienna, Faculty of Business, Economics and Statistics, Department of Business Administration, Chair of International Marketing, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

ARTICLE INFO

Article history:
First received on August 4, 2018 and was under review for 8 months
Available online 25 February 2020

Editor: Benedict Dellaert

Keywords:
Warmth
Competence
Stereotype content model
Construct measurement

A dynamic and growing body of marketing literature draws from the Stereotype Content Model's (SCM) dimensions, *warmth* and *competence*, to describe individuals' perception of various market-related stimuli (e.g., salespeople, employees, products) and predict their influence on subsequent behavior. However, the operationalization of these fundamental dimensions in empirical marketing research has been inconsistent and problematic, hindering comparability of findings across studies and limiting the overall usefulness of warmth–competence in relevant research. Using a mixed-method approach, the present paper scrutinizes the warmth–competence dimensions in order to validate and optimize their measurement. Through seven studies using different methodologies and metrics, we identify a set of trait items that *(a)* capture perceptions of warmth and competence consistently across a variety of impression formation settings, and *(b)* display strong psychometric properties. Our findings offer a parsimonious approach that promotes the consistent operationalization of warmth and competence across marketing studies and enables the integration of empirical findings, both of which are crucial requirements for meta-analytic investigations, replication studies, and integrative reviews.
© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The Stereotype Content Model (SCM; Fiske, Xu, Cuddy, & Glick, 1999; Fiske, Cuddy, Glick, & Xu, 2002) is probably the most widely used framework for studying the nature and consequences of stereotyping. According to the SCM, there are two fundamental dimensions, *warmth* and *competence*, which can systematically organize the content of people's stereotypical beliefs about "others" (Fiske et al., 2002; Kervyn, Yzerbyt, Demoulin, & Judd, 2008) and enable the prediction of distinct affective and behavioral responses (Cuddy, Fiske, & Glick, 2008; Fiske, Cuddy, & Glick, 2007). These dimensions are based on evolutionary theory and suggest that when people encounter others they are primarily interested in two things (Fiske et al., 2002, 2007): *(a)* identifying what the others' intent is, that is whether others have positive or negative intentions toward them (warmth), and *(b)* identifying how effectively others can pursue their intentions, that is how capable others are to either benefit or harm them (competence).

* Corresponding author.
E-mail addresses: georgios.halkias@univie.ac.at, (G. Halkias), adamantios.diamantopoulos@univie.ac.at. (A. Diamantopoulos).

https://doi.org/10.1016/j.ijresmar.2020.02.004
0167-8116/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Although different labels have been used in the literature (e.g., *communality* and *agency* (Conway, Pizzamiglio, & Mount, 1996) or *morality* and *competence* (Phalet & Poppe, 1997; Wojciszke, 1997), there is a high degree of consensus that warmth and competence can universally capture individuals' perception (Fiske, 2012, 2015, 2018; Kervyn et al., 2008). Indeed, through cultural, historical, and empirical insights, it emerges that these two dimensions transcend perceptual domains as well as levels of analysis and guide how people make sense of their social environment (Fiske, 2015, 2018). Overall, warmth and competence represent effective tools for analyzing people's perception of various social targets, ranging from entire nations and distinct social groups to specific individuals and their own self, or even the perception of non-human entities (Fiske, 2015, 2018; Fiske et al., 2007; Judd, James-Hawkins, Yzerbyt, & Kashima, 2005). Importantly, the dimensions of warmth and competence have been particularly influential to a growing body of research that permeates core marketing domains such as branding (Aaker, Garbinsky, & Vohs, 2012; Davvetas & Halkias, 2019; Kervyn, Fiske, & Malone, 2012), corporate/firm image (Aaker, Vohs, & Mogilner, 2010; Lepthien, Papies, Clement, & Melnyk, 2017), services (Habel, Alavi, & Pick, 2017; Kirmani, Hamilton, Thompson, & Lantzy, 2017), international marketing (Chen, Mathur, & Maheswaran, 2014; Diamantopoulos, Florack, Halkias, & Palcu, 2017), advertising (Zawisza & Pittard, 2015), and various consumer behavior phenomena (Dubois, Rucker, & Galinsky, 2016; Van der Lans, van Everdingen, & Melnyk, 2016; Zhou, Kim, Wang, & Aggarwal, 2018).

Notwithstanding the importance and immense penetration of the SCM in relevant literature,[1] there are some troublesome issues related to the operationalization of its dimensions. Specifically, and as will be further elaborated in the following sections, the original scale development process followed by Fiske et al. (1999) to generate multi-item measures of warmth and competence raises concerns about the content validity, dimensionality, and reliability of the chosen items (i.e., stereotypical traits). Such concerns are further vindicated by the lack of stability of the original items in subsequent empirical applications and their often arbitrary replacement with other (untested) items. Even studies within the same marketing domain employ differing operationalizations – either by means of experimentally manipulated factors or directly measured variables – often relying on indicators that have not been selected through systematic examination and validation procedures (e.g., Habel et al., 2017 vs. Wang, Mao, Li, & Liu, 2016 vs. Kirmani et al., 2017 in services; Kervyn et al., 2012 vs. Wu, Chen, & Dou, 2017 vs. Stokburger-Sauer, Ratneshwar, & Sen, 2012 in branding; Diamantopoulos et al., 2017 vs. Chen et al., 2014 in international marketing). These anomalies can lead to severe misinterpretations as to what these dimensions essentially represent and hinder the integration of empirical results since operationalizations supposedly capturing the *same* conceptual dimension often involve the manipulation and/or measurement of *different* (in)dependent variables. As a consequence, generalizability suffers and so does the theoretical and practical usefulness of warmth and competence as tools guiding the investigation of relevant marketing phenomena.

The present paper revisits the operationalization of warmth and competence and scrutinizes their content. In seven complementary studies involving a mixed-method approach, we undertake a rigorous examination of the SCM dimensions and identify a set of trait items that *(a)* capture perceptions of warmth and competence consistently across a variety of impression formation settings, and *(b)* display strong psychometric properties. The outcome of these efforts is a robust and parsimonious operationalization of warmth and competence that contributes to their consistent application across studies, thus enabling the integration of empirical results: a crucial issue in meta-analytic investigations, replication studies, and integrative reviews. From a methodological perspective, we also propose novel tasks and associated metrics for the assessment of content validity and item dimensionality during the initial stages of measure development. Finally, our findings corroborate the contribution of warmth–competence perceptions in predicting consumers' product and brand preferences.

In the next section, we provide evidence for the relevance of the warmth–competence dimensions in the marketing literature, highlighting several inconsistencies associated with their operationalization. Next, we revisit the process adopted by Fiske et al. (1999) to generate the original measures of warmth and competence and identify a number of critical issues related to item retention, sample type and size, and statistical procedures employed. This analysis is not intended as a criticism of the SCM per se but as a demonstration that, from a psychometric perspective, the original operationalization was problematic. We then apply diverse methodological procedures to assess the suitability of the warmth and competence trait items. Through this multi-method approach, we identify a set of items that are unambiguous, intuitive, homogeneous, and able to consistently capture individuals' perceptions across different judgment formation contexts. We conclude the paper by highlighting its implications, identifying relevant limitations and offering suggestions for future methodological work on the warmth–competence dimensions.

2. Perceptions of warmth and competence in marketing research

The dimensions composing the SCM were originally developed to explain intergroup perceptions and, in particular, social stereotypes: society's overgeneralized set of beliefs about the characteristics of the various social groups (Fiske et al., 2002). However, these dimensions have been extended well beyond the realm of social stereotyping and shown to be central in how people perceive their social environment in general (Fiske, 2018). Research evidence indicates that the warmth-by-competence space can effectively describe people's perception about very broad social classes (e.g., race), superordinate social categories (e.g., nationality), distinct societal groups (e.g., profession), more specific/narrow group subtypes (e.g., a female professor), personalities (e.g., introverts), other individuals, and their own self (see, Abele, Cuddy, Judd, & Yzerbyt, 2008; Cuddy et al., 2008, 2009; Fiske et al., 2007; Judd et al., 2005). Even non-human entities associated with perceived intent and agency, such as animals or brands, are mentally categorized

[1] The impact of the SCM in the scholarly literature is aptly illustrated by the number of citations of Fiske et al.'s (2002) seminal paper which, currently, approximates 5600 in Google Scholar (> 2415 Web of Science) [accessed October 2019].

Table 1
Marketing literature drawing (empirically and/or theoretically) from the universal dimensions of *warmth* and *competence*.

Author(s)	Journal	Application	Domain
Chang, Li, Yan, and Kumar (2019)	*Journal of Academy of Marketing Science*	Theoretical/empirical	Perception of brands
Zhou, Kim, and Wang (2019)	*Journal of Consumer Research*	Theoretical/empirical	Perception of money
Li, Chan, and Kim (2018)	*Journal of Consumer Research*	Theoretical/empirical	Perception of employees/salespeople/ service providers
Liu and Lin (2018)	*Journal of Consumer Psychology*	Theoretical/empirical	Perception of the self
Magnusson, Westjohn, and Siriani (2018)	*Journal of International Business Studies*	Theoretical/empirical	Perception of countries/brands
Kirmani et al. (2017)	*Journal of Marketing*	Theoretical/empirical	Perception of employees/salespeople/ service providers
Lepthien, Papiers, Clement, and Melnyk (2017)	*International Journal of Research in Marketing*	Theoretical/empirical	Perception of brands/firms
Habel et al. (2017)	*International Journal of Research in Marketing*	Theoretical/empirical	Perception of employees/salespeople/ service providers
MacInnis and Folkes (2017)	*Journal of Consumer Psychology*	Theoretical	Perception of brands/firms
Singh, Marinova, Singh, and Evans (2017)	*Journal of Academy of Marketing Science*	Theoretical	Perception of employees/salespeople/ service providers
Schultze, Gerlach and Rittich (2017)	*Journal of Behavioral Decision Making*	Theoretical/empirical	Perception of the self and advisors
Wang, Mao, Lin, and Liu (2017)	*Journal of Consumer Research*	Theoretical/empirical	Perception of employees/salespeople/ service providers
Gershon and Cryder (2017)	*Journal of Consumer Research*	Theoretical/empirical	Perception of brands/firms
Peter and Ponzi (2017)	*Journal of Advertising Research*	Theoretical/empirical	Perception of brands
Wu et al. (2017)	*Journal of Marketing Management*	Theoretical/empirical	Perception of brands/firms
Diamantopoulos et al. (2017)	*Journal of International Business Studies*	Theoretical/empirical	Perception of countries
Van der Lans et al. (2016)	*International Journal of Research in Marketing*	Theoretical/empirical	Perception of the self
Dubois et al. (2016)	*Journal of Consumer Research*	Theoretical/empirical	Perception of messages
Halkias et al. (2016)	*Journal of Business Research*	Theoretical/empirical	Perception of countries
Voliotis, Vlachos, and Epitropaki (2016)	*Frontiers in Psychology*	Theoretical/empirical	Perception of brands/firms
Antoneti and Maklan (2016)	*Psychology and Marketing*	Theoretical/empirical	Perception of brand users/consumers
Harris et al. (2016)	*Journal of Behavioral Decision Making*	Theoretical/empirical	(Employers') Perception of employees
Portal, Abratt, and Bendixen (2015)	*Business Horizon*	Theoretical	Perception of brands/firms
Zawisza and Pittard (2015)	*Basic and Applied Social Psychology*	Theoretical/empirical	Perception of advertising
Chen et al. (2014)	*Journal of Consumer Research*	Theoretical/empirical	Perception of countries
Kervyn, Chan, Malone, Korpusik, and Ybarra (2014)	*Social Cognition*	Theoretical/empirical	Perception of brands/firms
Scott, Mende, and Bolton (2013)	*Journal of Marketing Research*	Theoretical/empirical	Perception of employees/salespeople/ service providers
Aaker et al. (2012)	*Journal of Consumer Psychology*	Theoretical/empirical	Perception of brands/firms
Bennet and Hill (2012)	*Journal of Consumer Psychology*	Theoretical/empirical	Perception of brands/firms
Fournier and Alvarez (2012)	*Journal of Consumer Psychology*	Theoretical	Perception of brands/firms
MacInnis (2012)	*Journal of Consumer Psychology*	Theoretical	Perception of brands/firms
Maher and Carter (2011)	*International Marketing Review*	Theoretical/empirical	Perception of countries
Chattalas, Kramer, and Takada (2008)	*International Marketing Review*	Theoretical	Perception of countries

Note: The work listed hereby is by no means exhaustive and only serves illustrative purposes. Articles are presented in chronological order.

on the basis of warmth–competence (Fiske, 2015; Kervyn et al., 2012). As Dupree and Fiske (2017, p. 28) recently noted, "warmth and competence dimensions are universal dimensions of social perception that endure across stimuli, time, and place."

Marketing scholars have recently also started to utilize these dimensions in their research and, by now, a noticeable body of literature draws from warmth and competence in order to describe a wide range of market-related phenomena (see Table 1). For instance, researchers make use of the warmth–competence framework in order to investigate behavioral outcomes (e.g., attitude, adoption, charitable giving) based on how individual consumers perceive typical brand users (Antonetti & Maklan, 2016), themselves (Van der Lans et al., 2016), different product origins (Diamantopoulos et al., 2017), entire corporations (Aaker et al., 2010), individual brands (Davvetas & Halkias, 2019; Ivens, Leischnig, Muller, & Valta, 2015), or even inanimate objects imbued with anthropomorphic features (Zhou et al., 2018). The dimensions of warmth and competence have also been particularly influential in studying interpersonal perception both in terms of customer–employee encounters in the context of sales/services and in terms of employee–employee interactions within organizations (Harris, Lee, Thompson, & Kranton, 2016; Kirmani et al., 2017).

In summary, an increasing number of marketing studies theorize on the basis of these universal dimensions of perception and empirically operationalize them in experimental and survey-based research either as behavioral antecedents (causes), explanatory mechanisms (mediators), and/or measured responses (outcomes).

Despite the diffusion and growing importance of the warmth–competence dimensions in marketing research, their applications have been quite diverse and often incompatible. Table 2 offers an overview of recent marketing studies that document a number of discrepancies and problematic aspects in how warmth and competence are operationally approached.

Table 2
Differences/inconsistencies in the empirical operationalization of *warmth* and *competence* across marketing studies.

Article	Target/object of perception	Operationalization	Items/Indicators
Kervyn et al. (2012) *Journal of Consumer Psychology*	Brands/firms	Study 1: Experimental manipulation	Warmth: "*…with (without) public's best interest*" and "*…having (lacking) good intentions*" Competence: "*…(un)skilled and (in)effective at achieving its goals*" and "*…having (lacking) the ability to implement its intentions*"
		Manipulation check	Warmth *warm* *friendly* Competence *competent* *capable*
		Study 2: Direct measurement (independent variable)	Warmth *has good intentions* *has the public's best interest*▽ Competence *has ability to implement its intentions* *is skilled and effective achieving its goals*✝
Aaker et al. (2010) *Journal of Consumer Research*	Brands/firms	Study 1–3: Direct measurement (dependent variable)	Warmth *warm* *Kind* *generous*▽ Competence *competent* *effective*▽ *efficient*
Stokburger-Sauer et al. (2012) *International Journal of Research in Marketing*	Brands	Study 1: Direct measurement (independent variable)	Warmth⊗ *creates warm feelings*▽⊗ *is very loveable*▽⊗ *is emotional rather than rational*▽⊗
Ivens et al. (2015) *Psychology and Marketing*	Brands	Study 1: Direct measurement (mediating variable)	Warmth *good-hearted*▽ *warm* *with good intentions* Competence *effective*▽ *capable* *competent*
Habel et al. (2017) *International Journal of Research in Marketing*	employees/salespeople/ service providers	Study 1: Direct measurement (mediating variable)	Competence *well-prepared*▽ *proficient*▽ *capable* *competent*
		Study 2: Direct measurement (dependent variable)	Warmth *empathetic*▽ *friendly* Competence *well-prepared*▽ *competent*
		Study 3: Direct measurement (dependent variable)	Warmth *friendly* *warm* Competence *competent* *capable*
		Study 5–7: Direct measurement (mediating variable)	Warmth *warm* *friendly* *likeable* *empathetic*▽ Competence *well-prepared*▽ *proficient*▽ *capable* *competent*
Wang et al. (2016) *Journal of Consumer Research*	employees/salespeople/ service providers	Study 1a: Direct measurement (dependent variable)	Warmth *warm* *kind*

(continued on next page)

Table 2 *(continued)*

Article	Target/object of perception	Operationalization	Items/Indicators
			friendly *sincere* Competence *intelligent* *skillful* *capable* *competent*
		Study 2a–2b: Direct measurement (dependent variable)	Warmth *warm* *kind* *friendly* *approachable*▽ Competence *intelligent* *skillful* *capable* *competent*
Kirmani et al. (2017) *Journal of Marketing*	employees/salespeople/ service providers	Study 1: Direct measurement (dependent variable)	Two raters coded consumers' reviews in terms of warmth and competence mentions. Warmth coding scheme: *"Warmth attributes also relate to interactions with others but focus on the kindness and friendliness of these; these include friendliness, bedside manner, and customer service."* Competence coding scheme: *"Competence attributes enable people to efficiently attain their goals or obstruct the goal attainment and include diligence, level of education, efficiency, knowledge and thoroughness"*
		Study 2: Direct measurement (dependent variable)	Competence *not clever/clever*▽ *incompetent/competent* *not knowledgeable/knowledgeable*▽ *unskilled/skilled*
		Study 3: Direct measurement (dependent variable)	Warmth *cold/warm* *unfriendly/friendly* *unsociable/sociable*▽ *not nice/nice* Competence *not clever/clever*▽ *incompetent/competent* *not knowledgeable/knowledgeable*▽ *unskilled/skilled*
		Study 4–5: Direct measurement (dependent variable)	Competence *not clever/clever*▽ *incompetent/competent* *not knowledgeable/knowledgeable*▽ *unskilled/skilled*
Soliman and Buehler (2018) *Journal of Behavioral Decision Making*	employees / salespeople / service providers	Study 1, 2, and 4: Direct measurement (independent/control variable)	Warmth *cold/warm* *unlikeable/likeable* *shy/outgoing*▽ *introvert/extrovert*▽ Competence *not competent/competent* *stupid/smart*▽ *unintelligent/intelligent*
Van der Lans et al. (2016) *International Journal of Research in Marketing*	Self	Study 1: Direct measurement (independent/moderating variable)	Warmth *affectionate*▽ *careful*▽ *Caring*▽ *Genuine*▽ *Trustworthy* Competence *capable* *confident* *intelligent* *practical*▽ *successful*▽

Table 2 *(continued)*

Article	Target/object of perception	Operationalization	Items/Indicators
Magnusson, Westjohn, and Siriani (2019) *Journal of International Business Studies*	Brands/countries	Study 1A: Experimental manipulation	Warmth: *charming*▽, *warm*, and *romantic*▽ were used as indicators of a warm country (Spain). Competence: organized▽, disciplined▽, methodical▽, and focused▽ were used as indicators of a competent country (Germany).
		Study 2A: Experimental manipulation	Warmth: *fun*▽, *flirty*▽, and *playful*▽ were used as indicators of a warm brand. Competence: diligence▽ and meticulousness▽ were used as indicators of brand competent.
			Manipulation check scales used different items based on a (unspecified) subset of items based on Cuddy et al. (2009)
Chen et al. (2014) *Journal of Consumer Research*	Countries	Study 3–4: Experimental manipulation	Participants were asked to write (or read in S4) about a vivid and detailed travel experience in [COUNTRY] using three of out of five words.
			Warmth-positive (and negative in S3) condition: *nice, empathetic*▽, *considerate*▽, *honest*▽ and *warm-hearted*▽ *(cold, pessimistic*▽, *indifferent*▽, *dishonest*▽ and *hard-hearted*▽*)* Competence-positive (and negative in S3) condition: *helpful, competent, resourceful*▽, *efficient,* and *knowledgeable*▽ *(inexperienced*▽, *incompetent, inefficient, unprofessional*▽ and *amateur*▽*)*
		Manipulation check	Warmth *warm* Competence *competent*

Note: The work listed hereby is by no means exhaustive and only serves illustrative purposes.

+ The second item consists of two different indicators of competence, i.e. "skilled" and "effective." The instructions read, *"Please indicate how well the following statements describe [BRAND]:"* Thus, agreement with the second item could be driven by either one or both brand qualities. Note that "skillful" was included in Fiske et al. (2002) original pool of items tested, whereas "effective" was self-selected.

⊗ The study altogether mistreats the dimension of *warmth* as representing *affect*.

▽ Item/indicator not included in the original operationalization of *warmth* and *competence* by Fiske et al. (1999, 2002).

The absence of a robust measurement scheme as well as the inconsistencies associated with the original development and application of the SCM dimensions (Fiske et al., 1999, 2002), have resulted in arbitrary and unsubstantiated operationalizations of warmth and competence. What is more, the lack of a systematic empirical operationalization might also be (at least partly) responsible for generating conceptual confusion and leading to altogether inaccurate applications of the SCM dimensions. Indeed, there are studies that not only arbitrarily operationalize the corresponding dimensions but also conceptually misinterpret the dimension of warmth as representing *affect* (e.g., Chattalas et al., 2008; Stokburger-Sauer et al., 2012). Despite any ostensible analogies with affect and cognition, warmth and competence explicitly refer to *cognitive* components of people's judgments and should be viewed as antecedents of subsequent affect elicited (Cuddy et al., 2008; Fiske et al., 2002).

3. Revisiting the operationalization of the SCM dimensions

Based on a thorough triangulation of empirical psychological research that goes back more than 85 years, Fiske et al. (1999) derived a pool of adjective traits corresponding to warmth and competence. More specifically, drawing from the pioneer investigations on stereotyping by Katz and Braly (1933) – later replicated and extended by Gilbert (1951) and Karlins, Coffman, and Walters (1969) – as well as the work of Conway et al. (1996) on communality and agency, Fiske et al. (1999) juxtaposed and cross-examined approximately 100 items, to derive 27 adjective traits that best fit with their conceptualization of warmth and competence.[2] Unfortunately, despite the diligent approach followed by Fiske et al. (1999) to generate a suitable item pool, the assessment of the latter and subsequent operationalization of the warmth and competence dimensions was highly problematic.

Fiske et al.'s (1999) original operationalization was based on a sample 42 undergraduate students who were instructed to rate 17 different social groups along the 27 adjective traits. After performing 17 separate exploratory factor analyses (EFAs), the authors

[2] The conceptualization of warmth and competence by Fiske et al. (1999) was itself based on a long and solid research tradition in psychological literature (Allport, 1954; Asch, 1946; Bakan, 1966; Eagly, 1987; Wojciszke, 1994).

cross-examined item loadings on warmth and competence and finally selected 10 items (competence: *competent, intelligent, confident, competitive, independent*; warmth: *likable, sincere, good-natured, warm, tolerant*) that seemed to load consistently across most, but not all, target groups (i.e., between 11 and 14). The authors subsequently computed average scores using the 17 target groups "as if they were respondents" (Fiske et al., 1999, p. 478) and used the group-level scores to *(a)* assess the reliability of the warmth and competence scales, *(b)* calculate the correlation between warmth and competence, *(c)* compare the 17 groups in terms of their mean warmth and competence scores, and *(d)* cluster the 17 groups using warmth and competence as clustering variables.

While Fiske et al. (1999) openly acknowledge limitations due to the choice of target groups, use of only positive traits as descriptors of warmth/competence, and employment of student samples, other – and, in our view, much more important – problematic aspects of their analysis have not been identified. First, and perhaps most obvious, asking respondents to perform 459 assessments (27 items × 17 stimulus groups) is almost certain to generate fatigue effects and biased responding, thus, compromising validity (Podsakoff, MacKenzie, & Podsakoff, 2012; Shadish, Cook, & Campbell, 2002). Second, the sample size used in their factor analysis is extremely small both in absolute terms ($n = 42$) and in terms of the associated cases-to-variables ratio (42:27 = 1.55:1); even the most "liberal" guidelines for minimum sample sizes in factor analysis recommend an n of *at least* 100 or a 5:1 cases-to-variables ratio (MacCallum, Widaman, Zhang, & Hong, 1999; Velicer & Fava, 1998). Thus, it is not surprising that the factor solutions and item loading patterns were unstable and varied considerably across the 17 stimulus target groups. Unfortunately, Fiske et al.'s (1999) paper does not report the number of factors extracted in each of the 17 EFAs performed nor does it mention whether the 10 items finally selected consistently loaded on the first two factors or (at least for some solutions) on the additional (non-hypothesized) third or fourth factor. Third, Fiske et al. (1999) conducted reliability analysis for the warmth and competence scales *only* at the group level, that is, after averaging warmth and competence scores across respondents for each target group. This not only further reduces the already small sample size – as now the unit of analysis is the group and not the individual respondent, yielding an $n = 17$ – but also masks potential differences in the reliability of the scales for *each* individual group. Fifth, the same "aggregation" problem applies to the relationship between warmth and competence; while a negative *group-level* correlation is reported, this does not necessarily imply that warmth and competence will also negatively correlate *within* each group. Sixth, as was the case for the factor analysis, the sample size ($n = 17$) used in the cluster analysis is far too small. Based on recent methodological research, *"at least 70 times the number of variables"* (Dolnicar, Grün, Leisch, & Schmidt, 2014, p. 302; original emphasis) should be used to derive a stable cluster solution. Last, but by no means least, only first-generation statistical techniques (e.g., EFA and correlation analysis) were used by Fiske et al. (1999) to derive the warmth and competence scales. While informative at preliminary stages of scale development, such techniques need to be accompanied by more robust techniques (e.g., confirmatory factor analysis, structural equation models) that enable a comprehensive assessment of the psychometric properties of multi-item measures (Netemeyer, Bearden, & Sharma, 2003; Raykov & Marcoulides, 2011).

In subsequent work, Fiske et al. (2002) employed both a student ($n = 73$) and a non-student ($n = 38$) sample and asked respondents to evaluate 23 social groups along nine items, reflecting the dimensions of warmth (i.e., *sincere, good-natured, warm, tolerant*) and competence (i.e., *competent, confident, independent, competitive, intelligent*). However, although the initial 27 items in the original Fiske et al. (1999) study were reduced to nine, the total number of items included in the questionnaire came to 26, demanding a total of 598 answers per respondent and, once again, raising serious fatigue concerns. Also, in this empirical investigation (Study 1, Fiske et al., 2002), the trait "likeable" was excluded (without explanation) from the assessment of warmth, even though Study 1 in Fiske et al. (1999) included it in the final solution and Study 2 used it as the *sole* indicator of warmth. In a follow-up study (Study 2, Fiske et al., 2002), the authors significantly decreased demands on participants by dividing their sample ($n = 148$ students) across different target groups, but also changed the items reflecting warmth (*friendly*, *well-intentioned*, and *trustworthy* were added, whereas *tolerant* was dropped) and competence (*capable*, *efficient*, and *skillful* were added, whereas *competitive* and *independent* was dropped). These scale revisions were apparently undertaken "to fit more closely with common usage and to see whether warmth would differentiate more strongly among groups and correlate more strongly with its hypothesize predictor, competition" (Fiske et al., 2002, p. 890). However, neither was the source of the new items disclosed nor were reasons for dropping items from the original scales of Fiske et al. (1999) offered.

The practice of arbitrarily including/excluding items has been observed in several studies both in psychology (e.g., Caprariello, Cuddy, & Fiske, 2009; Meagher, 2017) and marketing research, with inconsistencies being observed even across the studies of the same article (see Table 2). The lack of a consistent and rigorously validated operationalization of the warmth and competence dimensions has resulted in a fragmented literature, inhibiting the integration of empirical results across different studies and posing serious risks to cumulative science (Bagozzi, 1980; Bainter & Bollen, 2014; Schimmack, 2010). Methodological literature has repeatedly warned that inconsistent and invalid operationalizations of theoretical constructs lead to methodological problems, misleading inferences, and interpretational confounding (Anderson & Gerbing, 1988; Bagozzi, 1980; Burt, 1973). Bearing in mind that the validity of a measure reflects "the degree to which a concept (term, variable, construct) achieved theoretical and empirical meaning within the overall structure of one's theory" (Bagozzi, 1980, p. 162), insufficiently validated measures hinder understanding of the structural and logical relationships of the construct(s) under investigation with other theoretical constructs, as they are inherently prone to greater meaning variations (Burt, 1973). As Maruyama (1997, p. 273) aptly puts it, poor measure validity "causes great problems in trying to interpret paths in a model because variables are not exactly what we think they are and are calling them."

To conclude, despite their theoretical elegance and pervasive presence, the universal dimensions of the SCM are still in need of a consistent, stable, and psychometrically sound measurement scheme. It is the search for such a scheme that is described in the remainder of this paper.

4. Research context and overview of studies

Finding a research context that serves the objectives of our investigation requires tracing back the original development of the SCM dimensions and identifying a central and socially salient form of stereotyping. To this end, people's perception of different countries and nationalities corresponds to a typical form of social judgment and stereotyping (Fiske, 2012, 2018; Glick et al., 2006; Kervyn et al., 2008; Phalet & Poppe, 1997). Psychological literature considers the identification of people's national origin as a fundamental social categorization process that is both profound and intuitive (Brewer, 1988; Lee & Fiske, 2006). As with all stereotypes, through the process of socialization, individuals develop a specific set of beliefs that reflect perceptions about the attributes of different countries and, consequently, nationalities (Phalet & Poppe, 1997). Consistent with the notion of stereotyping, such perceptions do not only apply to people, but also to every attitude object for which category membership can be established. For example, whenever a product is identified as coming from a certain country, stereotypical country-related beliefs intuitively transfer to the impression we form about it and generate inferences about the features and the nature of the product (Chen et al., 2014; Diamantopoulos et al., 2017). In a similar sense, the résumé of a job candidate or an exchange student might be perceived differently depending on the candidate's national origin (Kervyn et al., 2008). In short, the stereotypes individuals hold for different countries exert a strong influence on their judgments and behavioral intentions across a wide range of social interactions (Phalet & Poppe, 1997). Importantly, in today's turbulent socio-economic and political environment (e.g., the Eurozone crisis, the new wave of immigration within and to the EU, and the recent resurgence of nationalistic movements around the world) that puts national boundaries in the foreground and emphasizes in- versus out-group distinctions, nationality-based judgments reflect one of the most frequent and relevant manifestations of social categorization and, thus, a particularly fitting context for the purposes of the present research.

We conduct seven complementary studies to scrutinize the stereotypical traits associated with the SCM dimensions and identify those traits best suited to function as indicators of the warmth and competence dimensions of the SCM. Our investigation starts with the original 27 items employed by Fiske et al. (1999) as well as eight additional items used in their follow-up studies (e.g., Fiske et al., 2002; Fiske & Cuddy, 2006). The first four studies examine the item pool of 35 items using diverse methodological tasks and metrics and their findings are juxtaposed before any items are dropped from further analysis. This approach provides different, yet complementary, information for understanding the impression formation conditions under which an individual trait adjective is particularly (in)effective as a potential indicator of warmth or competence. The items surviving the initial screening stage are then thoroughly examined through the remaining studies. An overview of the studies is provided in Table 3 (see Web Appendix A1 for further details).

5. Study 1

The language, and in particular the adjectives, that people use can be harnessed to provide a window into their thoughts, feelings, and biases (Bhatia, 2017; Rocklage & Fazio, 2015). Hence, they are of paramount importance when seeking to understand others' perceptions of the world (Geeraerts, 2010; Tuggy, 1993). As already noted, the items (i.e., stereotypical traits) proposed to capture the two fundamental dimensions of the SCM have not been used consistently across studies and/or scholars. Such

Table 3
Overview of studies.

Study 1 Semantic screening task	Three expert coders assessed the content validity of stereotypical trait items by identifying lexical ambiguities.
Study 2 Unaided sorting task	Participants ($n = 15$, seven females, $M_{age} = 23.47$) performed an unprompted, intuitive sorting of stereotypical trait items. Resulting groupings, absolute frequencies, and co-occurrences among items were used to assess construct and item dimensionality.
Study 3 Guided allocation task	Participants ($n = 28$, 19 females, $M_{age} = 24.61$) performed a theory-driven allocation task of stereotypical trait items. Frequency of allocations and co-occurrences among items were used to assess item dimensionality and homogeneity.
Study 4 Between-subjects survey in a comparative judgment formation setting	Participants ($n = 136$, 80 females, $M_{age} = 30.48$) were exposed to multiple target pairs and performed mutually exclusive item allocations. Frequency of allocations and co-occurrences among items were used to assess item homogeneity and strength in discriminating between targets.
Study 5 Between-subjects survey across multiple targets	Participants ($n = 312$, 168 females, $M_{age} = 32.59$) were exposed to multiple targets and rated them along the dimensions of warmth and competence using the revised items emerged through studies 1 to 4. Exploratory factor analysis was used to assess dimensionality and internal consistency of the final measures.
Study 6 Between-subjects survey across multiple targets	Participants ($n = 304$, 157 females, $M_{age} = 32.93$) were exposed to multiple targets and rated them along the dimensions of warmth and competence. Confirmatory factor analysis and structural equation modeling was used to establish factor structure, partition total variance into trait, method and error components, and confirm the dimensionality, reliability, and convergent/discriminant validity of the measures. Study 6 also provides evidence of predictive validity using a simple nomological network.
Study 7a/b Between-subjects survey across multiple targets	Participants ($n_{7a} = 302$, 154 females, $M_{age} = 40.42$ and $n_{7b} = 334$, 170 females, $M_{age} = 34.19$) in two countries were exposed to the same targets and completed the same questionnaire. Measures of warmth and competence were investigated in the context of a more complex and theoretically elaborate nomological network. Multi-group invariance testing between the two country samples was performed and the predictive validity of the measures was established.

inconsistency might (at least partly) be explained by more or less subtle variations in the relationship between the signifier (i.e., adjective) and its denotation (i.e., what it stands for) that guide researchers in using one item over another depending on the research context at hand. For instance, the trait "competitive" included in the original item pool of the SCM (Fiske et al., 1999) can be taken as denoting "highly qualified" but also as being indicative of a "fighting and rivalrous disposition." Based on the conceptualization of the SCM, the former denotation would classify "competitive" as an indicator of competence (as originally intended), whereas the latter denotation would put it under the dimension of warmth. Furthermore, the same item can have either a positive or a negative valence based on the dimension it is ascribed to. For example, we might use "passive" to describe an indifferent and quiet individual who does not have negative intentions toward us (warmth) but also to indicate that an individual is rather idle and less capable of pursuing his/her intentions (competence, as originally intended).

From the above, it follows that ensuring content validity by establishing the semantic transparency of the adjective traits used to capture the warmth and competence dimensions is a crucial first step when seeking to operationalize these dimensions. With this in mind, in our first study, we concentrated on identifying items that are perceived to be polysemous or ambiguous and, hence, of questionable content validity.

5.1. Method

A semantic screening task with the initial 35 items was performed by three expert coders (specialists in cognitive linguistics with research expertise is lexical typology and semantics). Coders were told they have to assess the lexical ambiguity of a series of words that people typically use to describe how different social groups are perceived. After these general instructions, coders were presented with the list of adjective traits and were asked to classify each one them in a binary fashion as either 0 = *unambiguous* or 1 = *ambiguous*. No reference to the SCM or its dimensions was made at any point. To minimize possible carry-over effects, the order of item presentation was randomized across coders. At the end of the task, coders were debriefed to confirm that they were naïve to the SCM framework.

5.2. Analysis and results

Inter-coder agreement for the item classifications was 84.8% (Fleiss' Kappa coefficient = .77). Disagreements for seven inconsistent categorizations were resolved through discussion between the coders and the authors until consensus was reached. Five adjective traits that were coded as ambiguous by one of the three coders (*determined*, *gullible*, *confident*, *spineless*, and *well-intentioned*) as well as two items that were characterized as ambiguous by two coders (*authoritarian*, and *independent*) were all classified as ambiguous. This decision was taken to eliminate any possibility of – even rare – perceived polysemy. Three further items were also flagged as problematic (*tolerant*, *competitive*, and *passive*) as they were consistently categorized as ambiguous by all three coders. Overall, eight items were identified as not having a stable semantic content. This suggests that their meaning is contingent on the (transient) contextual specification and/or on the individual's idiosyncrasy and, therefore, they are less reliable in capturing aggregate, consensual beliefs (results of the semantic screening task are summarized in Web Appendix A2).

6. Study 2

The second study focused on the dimensionality of the SCM. In particular, the purpose of Study 2 was to explore *(a)* whether the two fundamental dimensions of warmth and competence *naturally* occur when people consider several stereotypical traits (construct dimensionality), and *(b)* whether the traits that supposedly capture warmth and competence respectively, tend to fall within their intended dimensions (item dimensionality). To this end, we employed an unaided sorting task whereby participants could freely group the 35 adjective traits, forming dimensions that best reflect their intuitive stereotypical beliefs. In line with Koch, Imhoff, Dotsch, Unkelbach, and Alves (2016), such data-driven tasks utilize individuals' naturalistic perception of semantic relatedness and membership of different descriptive terms without any a priori assumptions or implicit priming regarding their dimensionality. Importantly, prior research has found that sorting tasks (even when up to 60 items have to be sorted) are intuitive, produce less fatigue and boredom to participants, and provide reliable results (Blanchard, Aloise, & DeSarbo, 2017; Blanchard & Banerji, 2016).

6.1. Method

An unaided sorting task with 15 participants was conducted (seven females, $M_{age} = 23.47$, $SD = 4.22$). Respondents were first presented with a cover story introducing the idea of social categorization and then with the list of 35 trait items which they were asked to group based on their semantic relatedness. Consistent with our research context, respondents were informed that the traits provided are used to describe how society sees people coming from other countries (Cuddy et al., 2009). Next, respondents were asked to indicate if they were not familiar with any of the adjectives (i.e., traits) and to provide a brief definition for every one of them. No misunderstandings were identified. All respondents were familiar with the 35 traits and their lexical meaning. In a third step, respondents were asked to form distinct groups by putting together the traits they believe are semantically related and may fall under a dimension. No restrictions whatsoever were imposed on respondents. Respondents were free to use as many adjectives as they wanted and to form as many groups as they thought was appropriate. At the end of the sorting task,

respondents were also asked to identify the common theme for each of the dimensions and to provide a corresponding descriptive label.

6.2. Analysis and results

Completion time duration across participants ranged from 6.25 to 26.12 min ($M = 16.09$, $SD = 6.10$). The number of dimensions that emerged ranged from two to eight per respondent ($M = 4.60$, $SD = 1.62$) with four individuals forming six or more dimensions. Task completion times were not related to either the number or to the type of dimensions formed by participants who, on average, spent 3.08 min per dimension formed. With the exception of three individuals who did not use 7–13 items, respondents used at least 30 of the stereotypical traits provided. *Gullible* was the trait most often excluded (four respondents), followed by *intelligent*, *competitive*, *likeable*, and *well-intentioned* (three respondents).

In a first exploratory stage, we qualitatively analyzed the data by examining the items used to form the emerging dimensions as well as the labels given to the latter. To this end, we developed a coding scheme based on the original operationalization of the SCM (Fiske et al., 1999, 2002): if an emerging dimension consisted of *only* warmth-related or competence-related items (as originally proposed), it was coded as *WARM* or *COMP*, respectively. Adopting a more relaxed criterion in order to enable an initial identification of problematic items, we also coded dimensions mainly consisting of warmth (competence)-related items plus a *single* competence (warmth)-related item as $WARM_{tainted}$ ($COMP_{tainted}$). The descriptive labels respondents used to name these dimensions (e.g., interaction-oriented, friendly, performance-oriented, capable) corroborated our coding scheme. Finally, for those dimensions that could not be assigned to any of the above four categories, either a common underlying theme was identified and coded accordingly, otherwise the category was coded as *Mixed*.

Overall, 69 dimensions were formed by the 15 respondents, out of which 35 included either only warmth-related or only competence-related traits. Of the remaining 34 dimensions, nine were identified as $WARM_{tainted}$ and five as $COMP_{tainted}$. Further, only 10 dimensions contained negatively valenced items and were labeled accordingly by only eight out of 15 participants (e.g., *weak* or *deceitful*). The rest of the dimensions (labeled *Mixed*) could not be assigned to our coding scheme as no clear theme was discernible in their content. With regard to the dimensions pertaining to the SCM (i.e., *WARM*, *COMP*, $WARM_{tainted}$, and $COMP_{tainted}$), it appears that respondents do intuitively think along these lines. Indeed, 14 out of the 15 participants formed at least one category corresponding to the dimension of warmth and 12 formed categories corresponding to the dimension of competence. These results support the construct dimensionality of the SCM (see Web Appendix A3 for full details).

Additional insights about potentially problematic items, we gathered by examining trait "intrusions" in the $WARM_{tainted}$ and $COMP_{tainted}$ dimensions. This revealed that four competence (*spineless, competitive, gullible, confident*) and three warmth (*sincere, helpful, good-natured*) items were (incorrectly) grouped together with other warmth and competence items, respectively (see Table 4). Interestingly, all four competence items that intruded the warmth dimension had been identified as ambiguous in Study 1.

Table 4
Coefficients of relative substantive agreement (*crsa*) for warmth and competence items (Study 2).

Dimension			
Competence		Warmth	
	crsa		*crsa*
Independent	0.93	Friendly	0.96
Capable	0.92	Warm	0.95
Skillful	0.92	Understanding	0.94
Competent	0.92	Gentle	0.94
Intelligent	0.90	Trustworthy	0.93
Efficient	0.90	Kind	0.92
Industrious	0.87	Well-intentioned	0.92
Determined	0.83	Likable	0.88
Confident ($1 \times WARM_{tainted}$)	0.79	Tolerant	0.86
Authoritarian	0.79	Helpful ($1 \times COMP_{tainted}$)	0.85
Competitive ($2 \times WARM_{tainted}$)	0.71	Nice	0.84
Passive*	0.50	Good-natured ($1 \times COMP_{tainted}$)	0.83
Spineless* ($4 \times WARM_{tainted}$)	0.49	Irritable	0.72
Gullible* ($2 \times WARM_{tainted}$)	0.43	Hostile	0.67
		Sincere ($3 \times COMP_{tainted}$)	0.64
		Cold	0.61
		Greedy	0.59
		Arrogant	0.53
		Complaining	0.52
		Egotistic*	0.50
		Whiny*	0.45

Note. Parentheses indicate how many respondents grouped the item along with items uniformly belonging to the opposite dimension.
* *crsa* does not meet pre-specified criterion.

To provide a more robust assessment of item dimensionality against the original operationalization of warmth and competence, the co-occurrences among the stereotypical traits across all respondents were calculated. A co-occurrence matrix was developed whereby cell entries represent the number of times (n_{ij}) an item *i* co-occurs together with item *j*, with the maximum possible cell value being the total number of respondents (i.e., sample size). In order to quantify the dimensionality of the items in relation to the theoretical, a priori dimensions of warmth and competence, a modified version of Anderson and Gerbing's (1991) substantive validity coefficient (*svc*) was used. This measure, denoted as *coefficient of relative substantive agreement* (*crsa*), reflects the proportion of correct classifications (i.e., co-occurrences consistent with the original operationalization of the SCM) divided by the sum of correct plus wrong classifications. For each individual item, *crsa* was calculated as follows:

$$crsa = \frac{\frac{\sum_{k=1}^{k-1} nc}{k-1}}{\left(\frac{\sum_{k=1}^{k-1} nc}{k-1}\right) + \left(\frac{\sum_{k=1}^{k} nw}{k}\right)} \quad (1)$$

where *nc* denotes the number of times a given item co-occurs with items of the correct predefined dimension (e.g., co-occurrences of a given warmth item with other items of the warmth dimension), *nw* denotes the number of times a given item co-occurs with items of the wrong predefined dimension (e.g., co-occurrences of a given warmth item with items of the competence dimension) and *k* represents the number of items specified to belong to the respective (either the correct or the wrong) dimension. The range of this measure is between 0 and 1, indicating the extent of homogeneity of a given item with other items that theoretically belong together. Consistent with methodological literature dealing with similar metrics (see Anderson & Gerbing, 1991; Fuchs & Diamantopoulos, 2012; Lawshe, 1975), only values >.50 were considered acceptable. Results are summarized in Table 4.

Out of the four competence items previously classified into the $WARM_{tainted}$ category, only *gullible* and *spineless* did not exceed the acceptable threshold value, as did *passive*. All the remaining competence items, showed a satisfactory degree of homogeneity with average *crsa* .861 ($SD = .071$). The content of the warmth dimension proved to be more complicated. Two items failed to pass the pre-specified threshold but all three warmth items previously classified as $COMP_{tainted}$ had satisfactory *crsa* values. Overall, warmth traits with acceptable *crsa* values were considerably more variable, ranging from .522 to .956 ($M = .795$, $SD = .153$).

7. Study 3

To further investigate item dimensionality, in Study 3 we employed a guided allocation task with the intention to see whether – provided with the definitions of warmth and competence – respondents' allocation of item traits replicates the original conceptualization of the SCM and to assess the extent to which each individual item is systematically perceived to reflect warmth, competence, or none of these dimensions. In contrast to the unaided sorting task of Study 2, we now adopted a more theory-driven approach by providing participants with formal definitions of the two universal dimensions (Fiske et al., 2002) but allowing them to freely decide if an item falls or not under any of them. Thus, our third study enables us to identify whether an item is perceived to be diagnostic of the fundamental SCM dimensions in the first place and, further, how strongly it is associated with its intended dimension.

7.1. Method

Twenty-eight respondents participated in a guided allocation task (19 females, $M_{age} = 24.61$, $SD = 4.60$). Initially, they were presented with a brief cover story introducing the idea of social categorization and provided with the formal definitions of warmth and competence, drawn from Fiske et al. (2002). Subsequently, participants were given the list of the 35 stereotypical traits and were asked to allocate them into three mutually exclusive categories depending on whether they fell under the definition of *(a)* warmth (WARM), *(b)* competence (COMP), or *(c)* neither (OTHER) dimension. Participants were told that the list of traits is merely a sample randomly drawn from a larger pool of traits that can be used to describe how society sees people coming from other countries (Cuddy et al., 2009) and that the items in the list may or may not adhere to the definitions of warmth and competence. Hence, the instructions along with the presence of the third (OTHER) option ensured that participants had no a priori constraints in terms of items' (non)membership in the dimensions and were completely free to assign as many or as few items – from none (0) to all (35) – as they thought appropriate under warmth or competence.

7.2. Analysis and results

Participants' allocations were quantified by calculating the frequency with which each item was assigned to one of three categories (i.e., WARM, COMP, and OTHER). Next, we conducted one-sample χ^2-tests, contrasting the observed allocations against the expectation that each item has equal probabilities of being allocated to any one of the three categories. This was used as a screening step to identify *(a)* items that respondents tend to systematically consider irrelevant with respect to the two universal dimensions, and *(b)* items for which dimensional membership cannot be systematically established. If an item is frequently allocated to the category OTHER and the corresponding χ^2-test is significant, then this item should not be seen as relevant to the original

Table 5
One-sample χ^2-tests and coefficient of dimensionality strength results (Study 3).

Adjectives	WARM	COMP	NEITHER	A priori SCM dimension	χ^2	*p*-value	*cds*
Determined	3	20	5	COMP	18.50	0.000	0.61
Confident	4	19	5	COMP	15.07	0.001	0.54
Competitive	1	23	4	COMP	30.50	0.000	0.79
Independent	1	22	5	COMP	26.64	0.000	0.75
Capable	2	23	3	COMP	26.00	0.000	0.75
Industrious	4	22	2	COMP	26.00	0.000	0.64
Efficient	1	24	3	COMP	34.78	0.000	0.82
Competent	2	23	3	COMP	30.07	0.000	0.75
Intelligent	3	18	7	COMP	12.93	0.002	0.54
Skillful	1	25	2	COMP	39.50	0.000	0.86
Authoritarian▼	5	15	8	COMP	5.64	0.060	0.36
Passive▼	5	10	13	COMP	3.50	0.174	0.18
Gullible†	4	7	17	COMP	9.93	0.007	0.11
Spineless†	4	7	17	COMP	9.93	0.007	0.11
Warm	25	2	1	WARM	39.50	0.000	0.82
Likable	23	1	4	WARM	30.50	0.000	0.79
Sincere*	19	5	4	WARM	15.07	0.001	0.50
Good-natured	23	1	4	WARM	30.50	0.000	0.79
Egotistic*	16	2	10	WARM	10.57	0.005	0.50
Kind	27	1	0	WARM	24.14	0.000	0.93
Well-intentioned	20	2	6	WARM	19.14	0.000	0.64
Cold	17	1	10	WARM	13.78	0.001	0.57
Tolerant*	19	5	4	WARM	15.07	0.001	0.50
Helpful	22	3	3	WARM	25.78	0.000	0.68
Friendly	27	1	0	WARM	24.14	0.000	0.93
Nice	25	1	2	WARM	39.50	0.000	0.86
Arrogant▼	14	6	8	WARM	3.71	0.156	0.29
Greedy▼	11	7	10	WARM	0.93	0.629	0.14
Trustworthy▼	12	8	8	WARM	1.14	0.565	0.14
Gentle▼	15	0	13	WARM	0.14	0.705	0.54
Whiny†	10	1	17	WARM	13.78	0.001	0.32
Hostile▼	12	4	12	WARM	4.57	0.102	0.29
Understanding▼	15	8	5	WARM	5.64	0.060	0.25
Irritable†	11	1	16	WARM	12.50	0.002	0.36
Complaining†	9	1	18	WARM	15.50	0.000	0.29

Note. Column entries for WARM, COMP, and NEITHER represent the number of times a given item is allocated to the corresponding category. Expected frequencies set to be equal across categories (uniform distribution).
† Systematically classified as NEITHER.
▼ No significant different across categories.
* *cds* value does not meet pre-specified criterion.

conceptualization of the SCM dimensions. On the other hand, non-significant differences between observed and expected frequencies of item allocations (regardless of how these are actually distributed across the three categories), would be an indication of dimensional uncertainty, again rendering the item problematic.

Five adjectives traits were systematically classified as neither warmth nor competence items (denoted with "†" in Table 5). Moreover, for eight additional items (denoted with "▼" in Table 5) no significant differences across categories were observed implying inconsistent item dimensionality. Classification patterns also show that most of the negative items (e.g., *"egoistic," "cold," "authoritarian," "passive," "hostile," "greedy"*) were not systematically classified as "OTHER," indicating that there is no systematic (method) bias driving respondents to uniformly treat negative items as belonging to the same category. Overall, nine out of the 13 items that were found to be problematic belong to the dimension of warmth.

Having filtered out problematic items, we then focused on the 22 items that were systematically assigned to their intended dimension. We used a standardized measure to assess the extent to which an item reflects its a priori correct (as opposed to the wrong) dimension, taking into consideration the extent to which that item is perceived to be irrelevant altogether. The measure, denoted as the *coefficient of dimensionality strength* (*cds*), was calculated as follows:

$$cds = \left(\frac{\sqrt{\sum_{i=1}^{n_{cw}} (c_i - w_i)^2}}{n_{cw}} \right) \left(1 - \frac{n_x}{n_x + n_{cw}} \right) \quad (2)$$

where n_{cw} denotes the number of participants who allocated a given item to any one of the two SCM dimensions (i.e., assigned it to either WARM or COMP), n_x denotes the number of participants who allocated a given item to neither dimension (i.e., assigned it to OTHER), c_i represents the number of participants who allocated item i to the a priori correct dimension, and w_i indicates the number of participants who allocated item i to the a priori wrong dimension. The *cds* measure takes values between 0 and 1

and indicates how strongly an item conforms to its intended dimension, weighted by the extent to which it is perceived to be irrelevant to the SCM. A *cds* value of 0 implies that participants allocated a given item evenly to the dimensions of warmth and competence.[3] As more participants allocate an item to its correct dimension and, at the same time, fewer participants assign that item to OTHER, *cds* values will be closer to 1. As with the *crsa* (see Study 2), values >.50 were considered satisfactory.

The *cds* values in Table 5 corroborate the χ^2 findings with regard to the 13 problematic items previously identified, since they all fall well below the acceptable level (ranging between .11 and .36). Moreover, of the 22 items significantly more often allocated to their correct dimension, only 3 did not manage to exceed the minimum *cds* threshold.

8. Study 4

Research on social perception suggests that, even though perceptions of warmth and competence represent relatively stable stereotypical beliefs about a particular social target, they may vary depending on whether a single target is considered independently or in conjunction with others (Kervyn et al., 2008; Kervyn, Bergsieker, Grignard, & Yzerbyt, 2016). Ideally, the items capturing warmth and competence should be stable in both independent and comparative judgment formation settings. That is, trait items should not only be valid and reliable indicators of their corresponding dimension when an individual target is (separately) evaluated, but also effectively *discriminate* between different targets and be able to do so consistently. Thus, an item of a given SCM dimension should be able to allow individuals judge the *relative* strength of its underlying dimension between different targets/objects of perception. In line with this reasoning, Study 4 follows a comparative approach, whereby respondents simultaneously evaluate two stimulus targets, and investigates whether and how effectively each item can discriminate between them.

8.1. Method

A sample of 136 participants (80 females, $M_{age} = 30.48$, $SD = 11.19$) was recruited in a between-subjects study. Respondents were randomly exposed to one out of three different country pairs (i.e., *Spain/Sweden*, *Japan/France*, and *Germany/USA*) – which are characterized by a different profile in terms of warmth/competence (Cuddy et al., 2008; Diamantopoulos et al., 2017) – and completed a mutually exclusive allocation task. Similar to our previous studies, participants initially read a brief cover story about social categorization and were then presented with the 35 stereotypical traits. Participants were told that all of these traits can be used to describe how society views people from other countries and were asked to allocate each trait to *one* country of the stimulus pair based on how the majority in society would think (Fiske et al., 2002). This "either/or" approach avoided unreflective, cross allocation of items and allowed us to more effectively assess the strength of each item in differentiating between the stimulus targets. If an item belongs to the dimension of warmth (competence), it should be used, in principle, to describe the warmer (more competent) rather than the more competent (warmer) country target in a given stimulus pair. The order of appearance between the two countries as well as the order of trait presentation was counterbalanced to avoid any carry-over effects. As in the previous studies, no mention of the SCM and its dimensions was made during the task.

8.2. Analysis and results

To assess the homogeneity of the items, a co-occurrence matrix similar to that used in Study 2 was developed for each of the three country pairs and the *crsa* metric described in eq. 1 was utilized.[4] Given the more conservative nature of the present task, the criterion set for acceptable homogeneity specified that an item should have a *crsa* value >.50 in *at least* two of three country pairs. Furthermore, to examine whether items are effective in discriminating between different targets, we conducted χ^2-tests contrasting participants' actual allocations to the two countries against an a priori allocation based on equal probabilities. For the items that systematically loaded to one of the two target countries, we also utilized a simple measure (hereafter referred to as *coefficient of discrimination*) which quantified discrimination strength by calculating the proportion of the absolute difference between allocations to the two stimulus countries relative to the total number of possible allocations (i.e., sample size). The minimum requirement for discrimination required that an item systematically loads to one of the two targets in *at least* two of three country pairs with a higher (lower) coefficient of discrimination indicating that the item is more (less) effective in capturing nuanced perceptual differences involved in comparative contexts. Items should meet the criteria set above for both homogeneity and discrimination. Results are summarized in Table 6.

Analysis with regard to homogeneity showed that three competence items fail to produce satisfactory results. More specifically, *crsa* values for *spineless* were below the acceptable threshold across all different country pairs, while *confident* and *gullible* passed the threshold value in only one out of the three groups. In terms of the warmth dimension, eight traits had to be eliminated; *cold, greedy*, *sincere*, *trustworthy* and *hostile* demonstrated insufficient homogeneity across all country pairs, whereas *arrogant*, *tolerant*, and *egoistic* were found homogeneous in only one instance. Regarding discrimination effectiveness, χ^2-test results showed that seven competence and 10 warmth items did not meet the discrimination criteria specified. Notably, χ^2 values for *passive* and *spineless* (intended dimension: competence) as well as for *good-natured*, *hostile*, *tolerant*, and *understanding* (intended dimension:

[3] A value of 0 is also obtained if the item is irrelevant to the SCM dimensions – which, however, cannot be the case since such items were eliminated in the purification stage described earlier.

[4] Item co-occurrences for each country were collapsed for the three pairs.

Table 6
Item homogeneity and discrimination strength (Study 4).

	Homogeneity (crsa)	Coefficient of discrimination
Competence		
Capable	0.59	0.66
Competent	0.57	0.75
Competitive†	0.55	0.51
Determined†	0.57	0.28
Efficient	0.59	0.64
Independent†	0.56	0.35
Industrious	0.58	0.78
Intelligent	0.58	0.70
Skillful	0.57	0.44
Authoritarian†	0.57	0.39
Confident*†	0.50	0.23
Passive†	0.49	0.12
Gullible*	0.45	0.64
Spineless*†	0.47	0.21
Warmth		
Complaining	0.55	0.48
Friendly	0.55	0.51
Irritable	0.54	0.34
Kind	0.59	0.35
Likable	0.56	0.38
Nice	0.55	0.22
Warm	0.59	0.51
Whiny	0.56	0.53
Arrogant*	0.49	0.46
Cold*	0.41	0.49
Egoistic*†	0.48	0.27
Greedy*†	0.47	0.21
Hostile*†	0.46	0.17
Sincere*†	0.48	0.37
Tolerant*†	0.49	0.08
Trustworthy*	0.48	0.37
Understanding†	0.55	0.22
Well-intentioned†	0.54	0.18
Helpful†	0.52	0.24
Gentle†	0.52	0.25
Good-natured†	0.54	0.17

Note. Values aggregated using the weighted arithmetic mean to account for differences in the number of participants exposed to the three stimulus country pairs employed.
* Item does not meet the criterion for homogeneity set (i.e., *crsa* > .50 in *at least* two of three country pairs).
† Item does not meet the criterion for discrimination set (*coefficient of discrimination* > .50 in *at least* two of three country pairs).

warmth) were non-significant across all three country pairs. Overall, 14 items satisfied both the homogeneity and discrimination criteria.

To provide a comprehensive picture of the overall performance of each of the 35 items considered, we integrated the empirical results of the different tasks involved in the studies reported above. Each study was designed to examine item performance from a different methodological perspective. Therefore, best fitted items should be able to survive the scrutiny of all four studies and fulfil the threshold criteria specified therein. As Table 7 shows, a total of 11 items (five for warmth and six for competence) consistently fulfilled the necessary requirements.

9. Study 5

To explore the underlying dimensional structure of the 11 items that survived Studies 1–4, an exploratory factor analysis (EFA) was performed and the internal consistency of the derived dimensions was subsequently assessed.

9.1. Method

A sample of 321 participants (168 females, $M_{age} = 32.59$, $SD = 12.06$) was recruited in a between-subjects, web-based study conducted with an online consumer panel provider. Participants were randomly assigned to one of seven groups corresponding to six stimulus countries (*Japan*, *Spain, Sweden*, *USA*, *Italy*, *Germany*, and *France*) and completed a self-administered questionnaire. Similar to Study 4, employing multiple different targets in independent groups ensured sufficient variability with regard to the content of country stereotypes, while minimizing respondents' fatigue. Consistent with prior relevant research, participants' country/national stereotypes were assessed using the third-person technique, asking them to indicate how most people in their society see the stimulus country and its people (Cuddy et al., 2008; Fiske et al., 2002). Responses along the competence and warmth items

Table 7
Properties of the items satisfying the criteria set in all four studies.

	Study 1	Study 2	Study 3	Study 4	
	Semantic screening	Homogeneity (crsa)	Dimensionality (cds)	Homogeneity (crsa)	Coefficient of discrimination
Competence					
Capable	√	0.92	0.75	0.59	0.66
Competent	√	0.92	0.75	0.57	0.75
Efficient	√	0.90	0.82	0.59	0.64
Skillful	√	0.92	0.86	0.57	0.44
Industrious	√	0.87	0.64	0.58	0.78
Intelligent	√	0.90	0.54	0.58	0.70
Warmth					
Friendly	√	0.96	0.93	0.55	0.51
Kind	√	0.92	0.93	0.59	0.35
Likable	√	0.88	0.79	0.56	0.38
Nice	√	0.84	0.86	0.55	0.22
Warm	√	0.95	0.82	0.59	0.51

Note. All items were unanimously coded as *not ambiguous* in Study 1. No item has been cross-classified in Study 2. All items were systematically allocated to their a priori dimension in Study 3 (χ^2s ≥ 24.14, *ps* < .001). Items systematically loaded to a specific stimulus country in (at least) two out of the three country pairs in Study 4 (χ^2s ≥ 4.42, *ps* < .035). Aggregate values for homogeneity (*crsa*) and coefficient of discrimination are reported for Study 4.

were operationalized through seven-point scales anchored at *disagree/agree*. The order of the constructs and items within each construct was fully randomized across participants.

9.2. Analysis and results

The suitability of the data for factor analytic purposes was met (Bartlett's test of sphericity χ^2 (55) = 2739.92, $p < .001$, Kaiser-Meyer-Olkin measure of sampling adequacy = .91, and anti-image correlation matrix diagonals >.87) as were the sample size requirements (cases-to-variables ratio > 29:1; see MacCallum et al., 1999). The EFA results (using principal axis extraction and oblique (oblimin) rotation) produced a clear two-factor solution ($r = .34$), with the first factor (competence) explaining 52.24% of variance and the second factor (warmth) an additional 23.50% (cumulative variance 75.74%). The eigenvalues for the first two factors were 5.75 and 2.59, respectively, while those for the remaining factors extracted were ≤ .49. All items loaded consistently on their corresponding factor with factor loadings ranging between.81 and .86 for warmth items and .67 and .88 for competence items. The reliability and internal consistency of the derived dimensions was satisfactory with $\alpha_{warmth} = .93$ and $\alpha_{competence} = .92$, while McDonald's omega (ω) – a measure of reliability that is not tied to the a priori assumption of essential tau-equivalence (Hayes & Coutts, 2019) – was also satisfactory with $\omega = .92$ for both warmth and competence.

10. Study 6

Study 6 employed a new sample and used confirmatory factor analysis (CFA) and structural equation modeling (SEM) to provide a more comprehensive assessment of the factor structure and the psychometric properties (dimensionality, reliability, convergent/discriminant validity, common method bias) of the warmth and competence measures derived in Study 5. Moreover, Study 6 was used to provide preliminary evidence of predictive validity by placing the dimensions of warmth and competence in a simple, yet theoretically relevant, nomological network. To this end, we drew from international marketing literature and particularly from research on country-of-origin (COO) effects. COO has a long tradition of more than four decades in marketing research, with more than 550 published articles documenting that product preferences are significantly influenced by consumers' perceptions about the country from which products originate (Lu, Heslop, Thomas, & Kwan, 2016). What is more, recent COO studies have successfully applied the warmth–competence framework to show that country-related stereotypical perceptions predict consumers' preferences, thus offering established nomological relationships (Chen et al., 2014; Diamantopoulos et al., 2017; Halkias, Davvetas, & Diamantopoulos, 2016).

10.1. Method

A sample of 304 participants (157 females, $M_{age} = 32.93$, $SD = 11.80$) was recruited in a between-subjects, web-based study conducted with an online consumer panel provider. Participants were randomly assigned to one of seven groups as in study 5 and completed a self-administered questionnaire. The first block of the questionnaire captured perceptions about the stimulus countries and also assessed participants' level of familiarity with these countries using three seven-point items adapted from Schlosser (2006). The second block included three seven-point items adapted from (Leong et al., 2008) measuring general preferences toward products originating from the stimulus country and a marker variable capturing respondents' familiarity with social networking sites (*not at all familiar/very familiar*) to assess common method variance (CMV). A short section on demographics was

placed at the end of the questionnaire. The order of blocks, constructs within blocks, and items within constructs was randomized across participants.

10.2. Analysis and results

The 11 items based on the EFA in Study 5 were specified as reflective indicators of a two-factor measurement model (comprising warmth and competence as latent variables) and subjected to a confirmatory factor analysis (CFA) using LISREL 8.80. Prior to estimating the model, we assessed the distributional assumption of the data. Mardia's test of multivariate normality revealed significant deviations (Mardia's multivariate skewness = 12.53, $z = 10.95$, $p < .001$; Mardia's multivariate kurtosis = 181.34, $z = 11.40$, $p < .001$). Therefore, to adjust goodness-of-fit for bias due to multivariate non-normality, the Satorra-Bentler scaled chi-square statistic (S-Bχ^2) was employed (Satorra & Bentler, 2001; the same adjustment is applied in all subsequent studies).

Estimation of the CFA model resulted in poor initial fit (S-Bχ^2 (43) = 184.63; RMSEA = 0.105; CFI = 0.964; SRMR = 0.095) and alerted us to problematic items. More specifically, *industrious, intelligent* and *warm* were found to have high cross-factor loadings as well as highly correlated error variances both between them and with other items (mean modification indices per item ≥17.70). These three items were, therefore, excluded and a new CFA was performed. The new measurement model produced acceptable fit (S-Bχ^2 (19) = 52.57; RMSEA = 0.076; CFI = 0.984; SRMR = 0.062). Factor loadings, *t* values, Cronbach's alpha, composite reliability (CR), and average variance extracted (AVE) indicated a high level of reliability and convergent validity for the derived warmth and competence measures (Table 8).

To further test the factor structure and assess method influences, we *(a)* compared a one-factor measurement model against our hypothesized two-factor model, and *(b)* partitioned the total variance into trait, method, and error components. The CFA results revealed that the hypothesized two-factor model produced a much better fit (S-Bχ^2 (19) = 52.57) than the one-factor model (S-Bχ^2 (20) = 573.09). The difference between S-Bχ^2 values for nested models does not typically correspond to the chi-square distribution. Thus, simply subtracting the S-Bχ^2 statistic of the baseline model from the corresponding S-Bχ^2 of the more restrictive (nested) model yields an invalid statistic for testing hypotheses about differences in model fit (Bryant & Satorra, 2012). To overcome this problem, we followed Bryant and Sattora's (2012) procedure to derive a scaling correction factor which eventually allows the calculation of a scaled chi-square difference (*scaled*$\Delta\chi^2$) that produces accurate and valid comparisons. The resulting difference in fit between the two models was highly significant (*scaled*$\Delta\chi^2$ (1) = 444.46, $p < .001$), supporting the two-factor structure proposed. Moreover, introducing a method factor as an additional influence on all warmth and competence items and decomposing the total variance into trait, method, and error components showed that method bias was not a problem in the study (% of variance: $\text{trait}_{\text{warmth/competence}}$ = 69%/57%, $\text{method}_{\text{warmth/competence}}$ = 11%/7%, $\text{error}_{\text{warmth/competence}}$ = 19%/34%).

Regarding predictive validity, we estimated a simple structural model in which the country stereotype dimensions of warmth and competence were specified as latent variables acting as predictors of participants' product preferences while controlling for country familiarity. This model enables us to assess whether stereotypical perceptions of country warmth and competence contribute in predicting product preferences beyond established country-level influences. Discriminant validity among the constructs was assessed using the Fornell and Larcker (1981) criterion; for each latent variable the corresponding AVE was higher than its squared correlation with the other latent variables in the model (Table 9).

Table 8
Measurement properties of the model variables (Study 6).

Warmth:	α = .92, CR = .93 (.93), AVE = .76 (.77)
Most people in my society believe that the following attributes describe [target country and its people]	
Friendly	.86* (.86*)
Kind	.84* (.83*)
Likeable	.85* (.85*)
Nice	.93* (.93*)
Competence:	α = .87, CR = .87 (.87), AVE = .63 (.64)
Most people in my society believe that the following attributes describe [target country and its people]	
Capable	.88* (.88*)
Competent	.82* (.81*)
Efficient	.75* (.75*)
Skillful	.73* (.73*)
Product preferences:	α = .89, CR = .90, AVE = .75
Whenever possible, I prefer buying products from [target country].	.81*
I like the idea of owing products from [target country].	.92*
I would feel good buying products from [target country].	.85*
Country familiarity (control variable):	α = .95, CR = .95, AVE = .86
How well do you know [target country]. (*not at all/completely*)	.96*
How familiar are you with [target country]. (*not at all/completely*)	.94*
How knowledgeable of [target country] are you. (*not at all/completely*)	.87*

Note. Column entries are standardized factor loadings. α = Cronbach's alpha; CR = construct reliability; AVE = average variance extracted. The psychometric properties are based on the structural model subsequently estimated (values in parentheses refer to the two-factor measurement model).
* $p < .001$.

Table 9
Discriminant validity assessment (Study 6).

	Warmth	Competence	Product preferences	Country familiarity
Warmth	**.76**			
Competence	.15	**.63**		
Product preferences	.02	.09	**.75**	
Country familiarity	.00	.00	.10	**.86**

Note. Bold numbers on the diagonal show the AVE. Numbers on the off-diagonal are squared correlations between the constructs.

To ensure that CMV did not inflate the estimates of the model relationships, we followed Lindell and Whitney's (2001) marker variable approach and re-adjusted the zero-order correlations among the observed variables by partialling out the effects of the marker variable. The significance of the resulting partial correlations did not change, indicating that CMV is not a problem in our analysis. A good overall model fit was obtained (S-Bχ^2 (71) = 129.83; RMSEA = 0.052; CFI = 0.985; SRMR = 0.045), showing that both competence ($b = .36$, $p < .001$) and warmth ($b = .15$, $p < .01$) positively and significantly predict participants' product preferences, beyond any variance explained by their familiarity with these stimulus countries ($b = .29$, $p < .001$).

11. Study 7a/b

Our last study was explicitly designed to assess the cross-national invariance of the warmth and competence measures and test their predictive validity in the context of a theoretically more elaborate nomological network in two different countries. More specifically, the model tested draws on the *Behaviors from Intergroup Affect and Stereotypes* map (BIAS; Cuddy, Fiske, & Glick, 2007; Cuddy et al., 2008) – a framework that complements the SCM. According to the BIAS map, the two cognitive dimensions of warmth and competence predict distinct affective responses which, in turn, drive behavioral tendencies toward a stimulus target (Cuddy et al., 2007). In line with this, and also building on recent empirical findings (Diamantopoulos et al., 2017; Maher & Carter, 2011), we specified a model in which country warmth and competence are linked to the uniformly positive and negative country-level emotions of admiration and contempt, respectively (see Fig. 1). These emotions are, in turn, specified to impact brand affect and, through it, brand purchase intentions. The outcome variables are also simultaneously controlled for the influence of perceived product–country typicality (Spielmann, 2016) and brand familiarity (Diamantopoulos et al., 2017), both of which have been identified as influential factors of consumers' purchase intentions in the relevant literature. Importantly, the model tested transcends perceptual domains, as it links *country*-level responses to specific, *brand*-level outcomes (and not just 'products coming from… [country]' in general), while including important product/brand controls. From a theoretical perspective, this model specification thus offers a rather conservative and particularly relevant investigation of the predictive validity of warmth and competence.

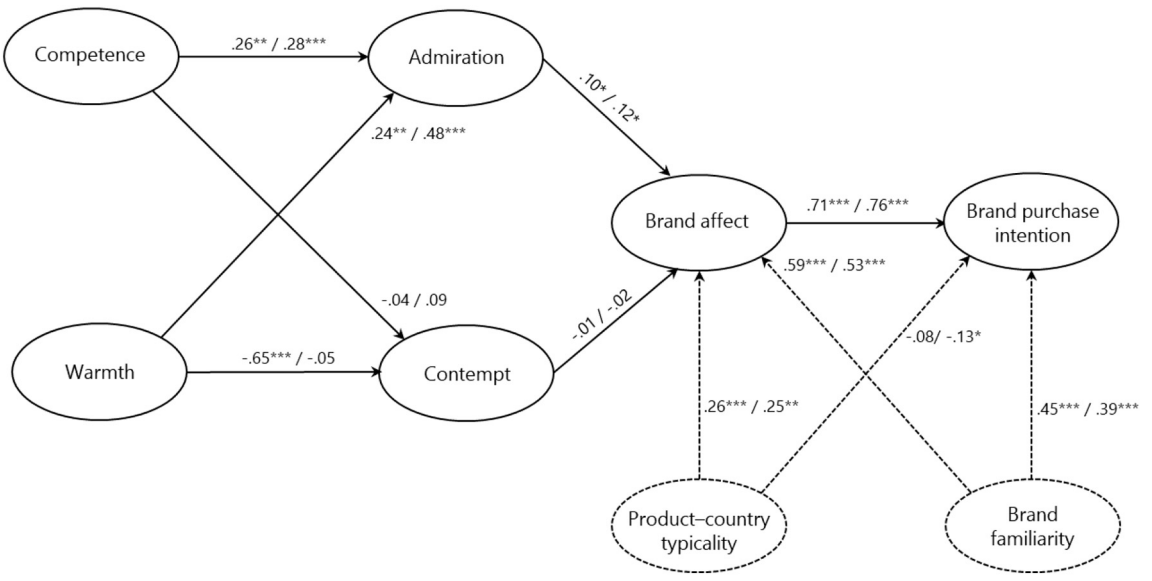

Fig. 1. Nomological validity results (Study 7a/Study 7b). *$p < .05$, **$p < .01$, ***$p < .001$.

11.1. Method

Study 7 involved an identical procedure administered in two different samples, Germany (Study 7a; $n = 302$, 154 females, $M_{age} = 40.42$, $SD = 12.77$) and the Netherlands (Study 7b; $n = 334$, 170 females, $M_{age} = 34.19$, $SD = 11.74$). Respondents were randomly assigned to one of seven between-subjects groups which corresponded to a different country/product category/brand combination (*Japan/TV/Sony*, *Spain/Clothes/Zara*, *Sweden/Furniture/IKEA*, *USA/Smartphones/Apple*, *France/Bottled water/Evian*, *Germany/Cars/Mercedes-Benz*, and *Italy/Perfumes/GiorgioArmani*). The first block of the questionnaire captured stereotypical country perceptions of warmth and competence and was identical to that of Studies 5 and 6, the only exception being that it also included single-item measures drawn from Cuddy et al. (2007) to assess country-level admiration and contempt. The second block measured brand affect, purchase intention, and brand familiarity as well as the level of perceived typicality between the stimulus countries and product categories. The questionnaire concluded with a marker variable (as in Study 6) and a short section on demographics. The order of blocks, constructs within blocks, and items within constructs was randomized across participants. Full details about the variables and their psychometric properties (factor loadings, Cronbach's alphas, CRs, AVEs, and discriminant validity) are provided in Tables 10 and 11.

11.2. Analysis and results

Prior to estimating the model in Fig. 1, we tested for measurement invariance of warmth and competence constructs using multi-group analysis (Table 12). Following the procedures outlined in Steenkamp and Baumgartner (1998), we first estimated a

Table 10
Measurement properties of model constructs (Study 7a/Study 7b).

Warmth:	α = .91/.92, CR = .91/.93, AVE = .72/.76
Most people in my society believe that the following attributes describe [target country and its people]	
Friendly	.89*/.86*
Kind	.82*/.88*
Likeable	.82*/.87*
Nice	.87*/.88*
Competence:	α = .88/.88, CR = .90/.89, AVE = .67/.67
Most people in my society believe that the following attributes describe [target country and its people]	
Capable	.86*/.84*
Competent	.91*/.83*
Efficient	.75*/.78*
Skillful	.73*/.82*
Country-level affect: To what extent do you think most people in your society have the following emotions toward *[target country and its people]*?	
Admiration (*not at all/very much*)	
Contempt (*not at all/very much*)	
Product–country typicality:	α = .96/.94, CR = .96/.94, AVE = .85/.81
This product category reflects [target country].	.87*/.86*
I associate this product category with [target country].	.95*/.92*
This product category makes me think of [target country].	.94*/.91*
There is a strong link between this product category and [target country].	.92*/.90*
Brand familiarity:	α = .95/.94, CR = .95/.94, AVE = .86/.85
How well do you know the brand? (*not at all/completely*)	.95*/.91*
How familiar are you with the brand? (*not at all/completely*)	.97*/.96*
How knowledgeable of the brand are you? (*not at all/completely*)	.85*/.88*
Brand affect:	α = .96/.96, CR = .96/.95, AVE = .88/.87
This brand makes me feel good	.93*/.95*
This brand makes me happy	.94*/.95*
This brand gives me pleasure	.94*/.90*
Brand purchase intention:	α = .98/.94, CR = .98/.94, AVE = .94/.84
I will likely buy this brand	.97*/.93*
The probability that I would consider buying this brand is high	.98*/.96*
I am willing to buy this brand	.95*/.86*

Note. Column entries are standardized factor loadings. α = Cronbach's alpha; CR = construct reliability; AVE = average variance extracted. All scale items were anchored *disagree/agree* unless otherwise indicated. Country-level affect employs single-item measures.

* $p < .001$.

Table 11
Discriminant validity assessment (Study 7a/Study 7b).

	Warmth	Competence	Admiration	Contempt	Product–country typicality	Brand familiarity	Brand affect	Purchase intention
Warmth	**.72/.76**							
Competence	.27/.10	**.67/.67**						
Admiration	.09/.25	.09/.12	**1.00**[⊗]					
Contempt	.27/.00	.09/.00	.03/.00	**1.00**[⊗]				
Product–country typicality	.02/.02	.09/.16	.01/.02	.01/.00	**.85/.81**			
Brand familiarity	.08/.01	.06/.06	.01/.01	.01/.00	.08/.06	**.86/.85**		
Brand affect	.02/.01	.06/.05	.02/.03	.01/.00	.17/.12	.47/.28	**.88/.87**	
Purchase intention	.02/.01	.05/.04	.01/.02	.01/.00	.08/.04	.59/.41	.67/.64	**.94/.84**

Note. Bold numbers on the diagonal show the AVE. Numbers on the off-diagonal represent the squared correlation between the constructs.
[⊗] Admiration and contempt are single-item constructs. Discriminant validity was determined by assessing the size of correlations and associated shared variance. Highest inter-correlations for admiration ($r_{study7a/7b}$ = .30/.50) and contempt ($r_{study7a/7b}$ = .51/.35) were small-to-medium with maximum shared variance ≤27%.

Table 12
Tests of measurement invariance for warmth and competence.

Measurement invariance	S-Bχ^2 (df)	Scaled $\Delta\chi^2$ (df)	RMSEA (90% CI)	CAIC	CFI	NNFI
Configural invariance *(equal form)*	51.49 (38)		0.033 (0.000–0.057)	304.97	0.997	0.996
Metric invariance *(equal factor loadings)*	50.88 (44)	4.65 (6)[†]	0.022 (0.000–0.045)	259.63	0.999	0.999
Scalar invariance *(equal factor loadings and intercepts)*	68.54 (50)	2.55 (6)[†]	0.034 (0.006–0.052)	351.84	0.996	0.996
Factor covariance invariance *(equal factor covariances)*	71.08 (51)	7.02 (1)[*]	0.035 (0.010–0.053)	346.92	0.996	0.995

[†] Not significant at the .05 level.
[*] $p < .01$.

multi-group CFA model without any restrictions in the parameters across the two samples (i.e., Germany and the Netherlands). Configural invariance was supported by a good model fit, indicating that the factor structure was stable in the two countries. Next, we tested for metric invariance by constraining all factor loadings to be equal in the two samples and also obtained a satisfactory model fit; importantly, the fit of the full metric invariance model was not significantly worse than that of the configural invariance model (*scaled*$\Delta\chi^2$ (6) = 4.65, *n.s.*). In a third step, we tested for scalar invariance by adding equality constraints on the item intercepts. The resulting model did not exhibit a significantly worse fit than the full metric invariance model (*scaled*$\Delta\chi^2$ (6) = 2.55, *n.s.*) suggesting full scalar invariance. Finally, we compared the scalar invariance model to a model where the covariance between warmth and competence was specified to be the same between the two countries. Despite returning a highly acceptable global fit, this restriction resulted in a significant deterioration of the fit as compared to the scalar invariance model (*scaled*$\Delta\chi^2$ (1) = 7.02, $p < .01$). Taken together, these results indicate that the measure function equivalently in both countries allowing inter-group comparisons of both relationships (i.e., effects) as well as latent variables means.[5]

Similar invariance tests were subsequently performed for all multi-item constructs included in the structural model in Fig. 1, resulting in partial metric invariance (see Web Appendix A4). Moreover, estimation of the proposed structural model resulted in good fit in both countries (Germany: S-Bχ^2 (215) = 305.62; RMSEA = 0.035; CFI = 0.992; SRMR = 0.061, Netherlands: S-Bχ^2 (215) = 318.70; RMSEA = 0.040; CFI = 0.991; SRMR = 0.053), providing additional evidence of the predictive validity of warmth and competence. Interestingly, the directionality and (non)significance of path coefficients are almost identical across the two countries (see Fig. 1). The results show that competence is significantly and positively related to feelings of admiration, whereas warmth is significantly associated with both decreased contempt and increased admiration. Moreover, positive country-related emotions spill over to the brand domain as indicated by a significant positive relationship between country-level admiration and brand-level affect which, subsequently, translates into higher brand purchase intentions. In both countries, these effects hold over and above any influence of product–country typicality and brand familiarity.

12. General discussion

Recent research based on the SCM and its fundamental dimensions of warmth and competence has rejuvenated and furthered the debate on the antecedents and consequences of social perception. Particularly in the marketing literature, an increasing number of studies draws from the universal dimensions that underlie the SCM in order to investigate a wide variety of phenomena that, among others, include understanding consumers' perceptions of brands (Kervyn et al., 2012), firms (Aaker et al., 2010), customer–employee interactions (Kirmani et al., 2017) and brand origin effects (Diamantopoulos et al., 2017). However, the

[5] The relationship between the warmth and competence dimension is higher in Germany (r = .51) than in the Netherlands (r = .30). The latent variable means, on the other hand, are practically identical across countries (warmth: $k_{Germany/Netherlands}$ = 5.19/5.13, competence: $k_{Germany/Netherlands}$ = 5.02/5.20).

original studies introducing the SCM (Fiske et al., 1999, 2002), although highly insightful, are subject to a number of methodological shortcomings and fail to provide robust evidence regarding the items (i.e., traits) that should be used to capture the warmth and competence dimensions. As a result, there is lack of consistency among marketing researchers – as well as among researchers in other disciplines – regarding the operationalization of these dimensions both as experimentally manipulated factors and as directly measured variables (e.g., Caprariello et al., 2009; Chen et al., 2014; Habel et al., 2017; Kervyn et al., 2008; Meagher, 2017; Stokburger-Sauer et al., 2012; Wang et al., 2016). Such inconsistency makes the integration of empirical applications highly problematic, inevitably leading to inconclusive results and altogether undermining the theoretical and practical relevance of these dimensions.

Already in their seminal paper, Fiske et al. (2002) highlighted the need for future studies employing more robust methodological approaches (e.g., more appropriate samples in terms of size and composition, techniques that undercut fatigue and carelessness bias, and cross validations) to refine and validate the warmth and competence scales. Yet, their call has not been heeded as evidenced by the lack of focused methodological investigations of the SCM dimensions. The present paper contributes in this direction by adopting a multi-method perspective in order to identify items that can consistently capture individuals' stereotypical perceptions in a variety of judgment formation settings. Using an extended pool of stereotypical traits drawn from the relevant literature, we applied different methodological approaches in seven complementary studies to derive a robust operationalization of the warmth–competence dimensions and offer strict tests of the dimensionality, reliability, convergent/discriminant validity, and predictive validity of the corresponding measurement scales.

Importantly, avoiding the shortcomings associated with the original operationalization of the SCM dimensions (Fiske et al., 1999), our studies employed sufficiently large, non-student samples, data collection instruments that do not wear out participants, and a variety of methodological tasks that allowed a more naturalistic investigation of the content of the warmth and competence dimensions. In this context, open (unaided) and closed (guided) sorting tasks are particularly useful in exploring perceptual structures because they provide researchers with flexibility in designing procedures that simulate different mental processes, are engaging and avoid participants' boredom/fatigue, and yield reliable results (Blanchard et al., 2017; Blanchard & Banerji, 2016). Indeed, through Studies 2, 3, and 4, we contribute to methodological literature by proposing novel quantitative indices (i.e., *crsa* and *cds*) for diverse classification and sorting tasks that can be utilized to assess the substantive validity and internal structure of constructs during the early stages of measure development (cf. Anderson & Gerbing, 1991).

Furthermore, application of second-generation statistical techniques in Studies 6 and 7a/b ensured that the psychometric properties of the final scales well-exceeded acceptable levels. The derived scales offer a parsimonious, less demanding to respondents (and, thus, easy to implement) approach to operationalize warmth and competence across different judgments formation settings. Consistent measurement of these dimensions (within and across disciplines) enables the integration and comparisons of empirical findings across studies which is essential in meta-analytic investigations for quantifying empirical inconsistencies and improving the estimation of the size of relevant effects.

In line with previous research suggesting that people are reluctant to report negatively framed beliefs but are less unwilling to report greater or lesser applicability of positive terms (Fiske et al., 2002), our findings strongly suggest that negative items perform poorly in capturing people's beliefs about socially sensitive issues, such as stereotyping. Importantly, susceptibility to social desirability and impression management influences not only undermines the validity of negatively valenced descriptors, but also suppresses ratings of low-to-medium values thus restricting variation and rendering them less effective measurement indicators (the latter point also resonates with the low metric scores that negative items achieved in Studies 2, 3, and 4). Nonetheless, it should be highlighted that although negative items might not be present in a measure, "negativity can (and does) come out in low ratings on positive traits" (Fiske et al., 1999, p. 487).

Although the main focus of this article has been methodological in nature, our findings also offer interesting theoretical insights. First, our results provide further evidence on the nature of stereotype content as captured by the SCM. Although our findings do not eliminate the possibility that further, less inclusive sub-dimensions might exist, they do confirm that warmth and competence are indeed core perceptual dimensions which occur regardless of the specific configuration of the impression formation setting. Second, our findings empirically support the relevance of the warmth dimension in the marketing context, which has been downplayed by prior research (Chen et al., 2014; Halkias et al., 2016). More specifically, under the implicit assumption that consumers mainly seek the best-performing products, international marketing literature has been traditionally treating COO information as an extrinsic cue signaling product quality and reliability (Maheswaran, 1994). In this context, country competence has been identified as the appropriate platform for companies wishing to develop COO-based strategies. In contrast to this notion, Studies 6 and 7 support the predictive strength of both dimensions, suggesting that country warmth and competence *(a)* directly predict consumers' general product preferences, and *(b)* indirectly lead to more favorable brand responses by increasing country-level emotions of admiration. Importantly, the nomological networks tested throughout our studies offer evidence to suggest that warmth and competence are constructs that can function independently from domain-specific knowledge related to the stimulus target at hand.

13. Limitations and future research

Previous studies based on the SCM have called for empirical investigations that focus on the operationalization of the warmth and competence dimensions in order to enable their effective application across social domains and research disciplines (Fiske et al., 2002; Kervyn et al., 2008). Our paper provides an answer to these calls, however, it does so by situating the empirical studies within the context of country/national stereotypes. Such stereotypes are socially prominent, relatively

stable within a given culture (Cuddy et al., 2009; Dovidio, 2013; Kervyn et al., 2008), and allow sufficient variability, thus serving well the purpose of our research. Nonetheless, further investigations using different marketing stimuli as targets are necessary to ensure that the derived warmth and competence scales are equally effective in capturing other forms of individuals' judgments. While one would not likely anticipate material differences in contexts involving human agents as stimulus targets (such as the perception of brand users, salespeople, and service employees), in cases where the target of perception refers to non-human entities or objects (such as a brand, a firm, or a retail environment), potential variations might be worth exploring as additional items might effectively capture warmth and competence and, thus, be used as proxies in empirical applications.

Moreover, the present paper is based on the core assumption that the SCM can be operationalized by means of a reflective measurement model, whereby the warmth and competence dimensions represent latent variables measured with reflective indicators that share a common theme and are, in principle, interchangeable (Bollen & Lennox, 1991). In this context, relevant literature has invariably approached these dimensions as multi-item reflective scales, without consideration of alternative measurement specifications. To this end, the SCM could be operationalized as a profile construct (Law, Wong, & Mobley, 1998), whereby different levels in each of the two universal dimensions would be combined to specify distinct *profiles* characterizing different stimulus targets. Future research taking this perspective could provide improved insights on mixed judgments and offer further evidence on the warmth–competence relationship.

Acknowledgment

The authors acknowledge the Austrian Science Fund (FWF) for supporting this study (research grant: P26740-G11). The authors are grateful to Prof. Hans Baumgartner, Penn State University, for his valuable insights on this project. The authors also wish to acknowledge the AE and reviewers of the *IJRM* for their constructive feedback throughout the review rounds.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijresmar.2020.02.004.

References

Aaker, J., Vohs, K. D., & Mogilner, C. (2010). Nonprofits are seen as warm and for-profits as competent: Firm stereotypes matter. *Journal of Consumer Research*, *37*(2), 224–237.
Aaker, J. L., Garbinsky, E. N., & Vohs, K. D. (2012). Cultivating admiration in brands: Warmth, competence, and landing in the "golden quadrant". *Journal of Consumer Psychology*, *22*(2), 191–194.
Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions of social judgment. *European Journal of Social Psychology*, *38*(7), 1063–1065.
Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423.
Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, *76*(5), 732–740.
Antonetti, P., & Maklan, S. (2016). Hippies, greenies, and tree huggers: How the "warmth" stereotype hinders the adoption of responsible brands. *Psychology & Marketing*, *33*(10), 796–813.
Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, *41*, 258–290.
Bagozzi, R. (1980). *Causal models in marketing*. New York: John Wiley and Sons Ltd.
Bainter, S. A., & Bollen, K. A. (2014). Interpretational confounding or confounded interpretations of causal indicators? *Measurement: Interdisciplinary Research & Perspectives*, *12*(4), 125–140.
Bakan, D. (1966). *The duality of human existence*. Chicago: Rand McNally.
Bhatia, S. (2017). The semantic representation of prejudice and stereotypes. *Cognition*, *164*, 46–60.
Blanchard, S. J., Aloise, D., & DeSarbo, W. S. (2017). Extracting summary piles from sorting task data. *Journal of Marketing Research*, *54*(3), 398–414.
Blanchard, S. J., & Banerji, I. (2016). Evidence-based recommendations for designing free-sorting experiments. *Behavior Research Methods*, *48*(4), 1318–1336.
Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305–314.
Brewer, M. B. (1988). A dual process model of impression formation. In R. S. Wyer, & T. K. Srull (Eds.), *Advances in social cognition* (pp. 177–182). New York: Psychology Press.
Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling*, *19*(3), 372–398.
Burt, R. S. (1973). Confirmatory factor-analytic structures and the theory construction process. *Sociological Methods & Research*, *2*(2), 131–190.
Caprariello, P. A., Cuddy, A. J. C., & Fiske, S. T. (2009). Social structure shapes cultural stereotypes and emotions: A causal test of the stereotype content model. *Group Processes & Intergroup Relations*, *12*(2), 147–155.
Chen, C. Y., Mathur, P., & Maheswaran, D. (2014). The effects of country-related affect on product evaluations. *Journal of Consumer Research*, *41*(4), 1033–1046.
Conway, M., Pizzamiglio, M. T., & Mount, L. (1996). Status, communality, and agency: Implications for stereotypes of gender and other groups. *Journal of Personality and Social Psychology*, *71*(1), 25–38.
Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*(4), 631–648.
Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 61–149.
Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J. -P., ... Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, *48*(1), 1–33.
Davvetas, V., & Halkias, G. (2019). Global and local brand stereotypes: Formation, content transfer, and impact. *International Marketing Review*, *36*(5), 675–701.
Diamantopoulos, A., Florack, A., Halkias, G., & Palcu, J. (2017). Explicit vs. implicit country stereotypes as predictors of product preferences: Insights from the stereotype content model. *Journal of International Business Studies*, *48*(8), 1023–1036.
Dolnicar, S., Grün, B., Leisch, F., & Schmidt, K. (2014). Required sample sizes for data-driven market segmentation analyses in tourism. *Journal of Travel Research*, *53*(3), 296–306.
Dovidio, J. F. (2013). Bridging intragroup processes and intergroup relations: Needing the twain to meet. *British Journal of Social Psychology*, *52*(1), 1–24.

Dubois, D., Rucker, D. D., & Galinsky, A. D. (2016). Dynamics of communicator and audience power: The persuasiveness of competence versus warmth. *Journal of Consumer Research*, *43*(1), 68–85.
Dupree, C., & Fiske, S. (2017). Universal dimensions of social signals: Warmth and competence. In J. Burgoon, N. Magnenat-Thalmann, M. Pantic, & A. Vinciarelli (Eds.), *Social Signal Processing* (pp. 23–33). Cambridge: Cambridge University Press.
Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation.* Hillsdale, NJ: Erlbaum.
Fiske, S. T. (2012). Journey to the edges: Social structures and neural maps of inter-group processes. *British Journal of Social Psychology*, *51*(1), 1–12.
Fiske, S. T. (2015). Intergroup biases: A focus on stereotype content. *Current Opinion in Behavioral Sciences*, *3*, 45–50.
Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*(2), 67–73.
Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.
Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902.
Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis) respecting versus (dis) liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, *55*(3), 473–489.
Fornell, D., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.
Fuchs, C., & Diamantopoulos, A. (2012). Positioning bases' influence on product similarity perceptions. *Quantitative marketing and marketing management* (pp. 325–351). Wiesbaden: Gabler Verlag.
Geeraerts, D. (2010). *Theories of lexical semantics.* Oxford University Press Inc.
Gilbert, G. M. (1951). Stereotype persistence and change among college students. *Journal of Personality and Social Psychology*, *46*(2), 245–254.
Glick, P., Fiske, S. T., Abrams, D., Dardenne, B., Ferreira, M. C., Gonzalez, R., ... Yzerbyt, V. (2006). Anti-American sentiment and America's perceived intent to dominate: An 11-nation study. *Basic and Applied Social Psychology*, *28*(4), 363–373.
Habel, J., Alavi, S., & Pick, D. (2017). When serving customers includes correcting them: Understanding the ambivalent effects of enforcing service rules. *International Journal of Research in Marketing*, *34*(4), 919–941.
Halkias, G., Davvetas, V., & Diamantopoulos, A. (2016). The interplay between country stereotypes and perceived brand globalness/localness as drivers of brand preference. *Journal of Business Research*, *69*(5), 3621–3628.
Harris, L., Lee, V. K., Thompson, E. H., & Kranton, R. (2016). Exploring the generalization process from past behavior to predicting future behavior. *Journal of Behavioral Decision Making*, *29*(4), 419–436.
Hayes, A. F., & Coutts, J. J. (2019). Use omega rather than Cronbach's alpha for reliability. But… *Communication Methods and Measures,* (forthcoming).
Ivens, B. S., Leischnig, A., Muller, B., & Valta, K. (2015). On the role of brand stereotypes in shaping consumer response toward brands: An empirical examination of direct and mediating effects of warmth and competence. *Psychology & Marketing*, *32*(8), 808–820.
Judd, C. M., James-Hawkins, L., Yzerbyt, V. Y., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology*, *89*, 899–913.
Karlins, M., Coffman, T. L., & Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, *13*(1), 1–16.
Katz, D., & Braly, K. W. (1933). Racial stereotypes of 100 college students. *Journal of Abnormal and Social Psychology*, *28*(3), 280–290.
Kervyn, N., Bergsieker, H. B., Grignard, F., & Yzerbyt, V. Y. (2016). An advantage of appearing mean or lazy: Amplified impressions of competence or warmth after mixed descriptions. *Journal of Experimental Social Psychology*, *62*, 17–23.
Kervyn, N., Fiske, S. T., & Malone, C. (2012). Brands as intentional agents framework: How perceived intentions and ability can map brand perception. *Journal of Consumer Psychology*, *22*(2), 166–176.
Kervyn, N., Yzerbyt, V. Y., Demoulin, S., & Judd, C. M. (2008). Competence and warmth in context: The compensatory nature of stereotypic views of national groups. *European Review of Social Psychology*, *21*, 155–187.
Kirmani, A., Hamilton, R. W., Thompson, D. V., & Lantzy, S. (2017). Doing well versus doing good: The differential effect of underdog positioning on moral and competent service providers. *Journal of Marketing*, *81*(1), 103–117.
Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The ABC of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of Personality and Social Psychology*, *110*(5), 675–709.
Law, K. S., Wong, C. S., & Mobley, W. M. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review*, *23*(4), 741–755.
Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575.
Lee, T. L., & Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: Immigrants in the stereotype content model. *International Journal of Intercultural Relations*, *30*(6), 751–768.
Leong, S. M., Cote, J. A., Ang, S. H., Tan, S. J., Jung, K., Kau, A. K., & Pornpitakpan, C. (2008). Understanding consumer animosity in an international crisis: Nature, antecedents, and consequences. *Journal of International Business Studies*, *39*(6), 996–1009.
Lepthien, A., Papies, D., Clement, M., & Melnyk, V. (2017). The ugly side of customer management–consumer reactions to firm-initiated contract terminations. *International Journal of Research in Marketing*, *34*(4), 829–850.
Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, *86*(1), 114–121.
Lu, I. R., Heslop, L. A., Thomas, D. R., & Kwan, E. (2016). An examination of the status and evolution of country image research. *International Marketing Review*, *33*(6), 825–850.
MacCallum, R., Widaman, K., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99.
Maher, A. A., & Carter, L. L. (2011). The affective and cognitive components of country image: Perceptions of American products in Kuwait. *International Marketing Review*, *28*(6), 559–580.
Maheswaran, D. (1994). Country of origin as a stereotype: Effects of consumer expertise and attribute strength on product evaluations. *Journal of Consumer Research*, *21*(2), 354–365.
Maruyama, G. (1997). *Basics of structural equation modeling.* Thousand Oakes: Sage Publications.
Meagher, B. R. (2017). Judging the gender of the inanimate: Benevolent sexism and gender stereotypes guide impressions of physical objects. *British Journal of Social Psychology*, *56*(3), 537–560.
Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications.* Thousand Oakes: Sage Publications.
Phalet, K., & Poppe, E. (1997). Competence and morality dimensions of national and ethnic stereotypes: A study in six eastern-European countries. *European Journal of Social Psychology*, *27*(6), 703–723.
Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539–569.
Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory.* New York: Routledge.
Rocklage, M. D., & Fazio, R. H. (2015). The evaluative lexicon: Adjective use as a means of assessing and distinguishing attitude valence, extremity, and emotionality. *Journal of Experimental Social Psychology*, *56*, 214–227.
Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514.
Schimmack, U. (2010). What multi-method data tell us about construct validity. *European Journal of Personality*, *24*(3), 241–257.
Schlosser, A. E. (2006). Learning through virtual product experience: The role of imagery on true versus false memories. *Journal of Consumer Research*, *33*(3), 377–383.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* New York: Houghton Mifflin.
Spielmann, N. (2016). Is it all or nothing? Testing schema congruity and typicality for products with country origin. *Journal of Business Research*, *69*(3), 1130–1137.
Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–90.
Stokburger-Sauer, N., Ratneshwar, S., & Sen, S. (2012). Drivers of consumer–brand identification. *International Journal of Research in Marketing*, *29*(4), 406–418.
Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics*, *4*(3), 273–290.

Van der Lans, R., van Everdingen, Y., & Melnyk, V. (2016). What to stress, to whom and where? A cross-country investigation of the effects of perceived brand benefits on buying intentions. *International Journal of Research in Marketing*, *33*(4), 924–943.
Velicer, W. F., & Fava, J. L. (1998). Affects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*(2), 231.
Wang, Z., Mao, H., Li, Y. J., & Liu, F. (2016). Smile big or not? Effects of smile intensity on perceptions of warmth and competence. *Journal of Consumer Research*, *43*(5), 787–805.
Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology.*, *67*(2), 222–232.
Wojciszke, B. (1997). Parallels between competence- versus morality-related traits and individualistic versus collectivistic values. *European Journal of Social Psychology*, *27*(3), 245–256.
Wu, J., Chen, J., & Dou, W. (2017). The internet of things and interaction style: The effect of smart interaction on brand attachment. *Journal of Marketing Management*, *33*(1–2), 61–75.
Zawisza, M., & Pittard, C. (2015). When do warmth and competence sell best? The "golden quadrant" shifts as a function of congruity with the product type, targets' individual differences, and advertising appeal type. *Basic and Applied Social Psychology*, *37*(2), 131–141.
Zhou, X., Kim, S., Wang, L., & Aggarwal, P. (2018). Money helps when money feels: Money anthropomorphism increases charitable giving. *Journal of Consumer Research*, *45*(5), 953–972.