Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

The Influence of Ambient Temperature on Social Perception and Social Behavior

Jan S. Krause, Gerrit Nanninga, Patrick Ring, Ulrich Schmidt, Daniel Schunk

May 21, 2020

Discussion paper number 2013

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
wiwi.uni-mainz.de

Contact Details:

Jan S. Krause
Kiel Institute for the World Economy
University of Kiel
24118 Kiel, Germany
jan.krause@economics.uni-kiel.de

Gerrit Nanninga
Department of Psychology
University of Kiel
24118 Kiel, Germany
gerritnanninga@gmx.de

Patrick Ring
Kiel Institute for the World Economy
University of Kiel
24118 Kiel, Germany
patrick.ring@ifw-kiel.de

Ulrich Schmidt
Kiel Institute for the World Economy
University of Kiel
24118 Kiel, Germany
ulrich.schmidt@ifw-kiel.de

Daniel Schunk
Chair of Public and Behavioral Economic
Gutenberg School of Management and Economics
University of Mainz
Jakob-Welder-Weg 9
55128 Mainz, Germany
daniel.schunk@uni-mainz.de

The Influence of Ambient Temperature on Social Perception and Social Behavior

Jan S. Krause[1,2], Gerrit Nanninga[3], Patrick Ring[1], Ulrich Schmidt[1,2,4] and Daniel Schunk[5*]

Abstract

The literature suggests that human perception and behavior vary with physical temperature. We provide an experimental test of how different ambient temperature conditions impact social behavior and social perception: Subjects went through a series of tasks measuring various aspects of social behavior and perception under three temperature conditions (cold vs. optimal vs. warm). Despite well-established findings on temperature effects, our data suggest that physical temperature has no relevant influence on social behavior and social perception. We corroborate our finding of a null effect by the use of equivalence testing and provide a discussion in the light of recent failed replication attempts in this field of research.

JEL classification: C90, D01, D90, D91

Keywords: social perception, ambient temperature, social preference, equivalence testing, cooperation, warmth

---

1. Kiel Institute for the World Economy, Kiel, Germany

2. Department of Economics, University of Kiel, Kiel, Germany

3. Department of Psychology, University of Kiel, Kiel, Germany

4. Department of Economics and Econometrics, University of Johannesburg, Johannesburg, South Africa

5. Department of Economics, University of Mainz, Mainz, Germany

* Corresponding author: Johannes Gutenberg-Universität Mainz, Fachbereich Rechts- und Wirtschaftswissenschaften, Lehrstuhl Public and Behavioral Economics, Jakob-Welder-Weg 4, D-55128 Mainz

Introduction

Thermoregulation is one of the most important functions of the human body (Charkoudian, 2003). Over time, evolutionary processes profoundly shaped mankind's behavior and cognition so that many human characteristics originate from adjusting to external temperature (Wheeler, 1984; Parsons, 2014). It is argued that humankind regressed hair growth and developed upright walking through evolutionary processes to facilitate the dissipation of excess body heat (Wheeler, 1984). Such adaptations were crucial for survival, since a 4°C rise from normal body temperature is fatal (Bouchama & Knochel, 2002). Consequently, thermoregulation deeply shaped human nature, society and culture. Temperature influences primal behavior like shadow seeking when it is hot or seeking warmth by huddling with conspecifics (Ebensperger, 2001). Humans consume food, clothe themselves and build houses to maintain a stable body temperature (Parsons, 2014). Oak tree ring analysis revealed a significant role of temperature changes to the major mass migrations in central Europe in the last 2500 years (Büntgen, Tegel, Nicolussi, McCormick, Frank, Trouet, ... & Luterbacher, 2011). Further, thermal stress has always been and still is a relevant factor in military operations (Goldman, 2001).

Neoclassical economic models do not take temperature into account. However, due to their economic relevance, economists have investigated effects of temperature empirically. Evidence linked high temperatures to lower stock market returns (Cao & Wei 2005), lower economic growth (Dell, Jones & Olken, 2012), less economic production (Burke, Hsiang & Miguel, 2015a) and hence a higher risk of conflict (Burke, Hsiang & Miguel, 2015b). Temperature also influences risk preferences (Wang, 2017) and court decisions (Heyes & Seberian, 2019). Finally, the relevance of studying the link between temperature and human behavior is obvious in the light (or shadow) of the predicted climate change resulting in a global rise in temperature and consequently an increased frequency and magnitude of heat waves (Hoegh-Guldberg, Jacob, Taylor, Bindi, Brown, Camilloni, ... & Guiot, 2018).

In contrast to the literature in psychology, the literature in economics does not offer a coherent framework explaining the empirically observed effects of temperature on human behavior. We therefore now review the psychological literature on temperature and behavior, and derive our hypotheses from this review. Temperature effects are an important topic in the psychological literature. It has been shown that external temperatures strongly influence and shape human psychology and behavior. Aggression and conflict are promoted through heat (Anderson, 1989; 2001; DeWall, Anderson & Bushman, 2011) and high temperatures impair mood (Keller, Fredrickson, Ybarra, Côté, Johnson, Mikels, . . . Wager, 2005), sleep (Okamoto-Mizuno, Mizuno, Michie, Maeda &

Iizuka, 1999) and health (Ormandy & Ezratty, 2012, Parsons, 2014). There is evidence that heat could also lead to more suicidal behavior (Holopainen, Helama & Partonen, 2013; Fountoulakis, Savopoulos, Zannis, Apostolopoulou, Fountoukidis, Kakaletsis, ... & Pompili, 2016). Multiple broad-based meta-studies suggest that extremely high and low temperatures negatively affect cognitive performance pointing into the direction that performance follows an inverse-U shaped curve with an optimum at medium temperatures (Pilcher, Nadler & Busch's, 2002; Hancock, Ross, & Szalma, 2007; Yeganeh, Reichard, McCoy, Bulbul, & Jazizadeh, 2018). Results from previous studies indicate negative emotional and behavioral responses to extreme temperatures. Importantly, there is evidence associating warmth with prosocial behavior. Asch (1949) claimed that the words "cold" and "warm" dramatically change the impression individuals form about others. Furthermore, there is evidence that thermal factors also influence social perceptions indicated by specific language use, i.e., idioms like "showing the cold shoulder" or "having warm feelings for someone" (Ijzerman & Semin, 2009). Metaphorical phrases of this kind appear universal across cultures (Lakoff & Johnson, 1980; Landau, Meier & Keefer, 2010). In our literature review we identified three main constructs positively impacted by warmth, social warmth, social distance and empathy which we briefly introduce in the following.

There is extensive psychological research about the connection between temperature and social warmth. Social warmth is one factor of a universal two-factorial person perception model: Socially interacting organisms must determine whether an opponent is "friend" or "foe" and to what extent an opponent could be harmful or helpful (referred to as ability; Fiske, Cuddy & Glick, 2007; Abele & Wojciszke, 2007). Based on Grounded Cognition Theories (Barsalou, 2008) there is prominent evidence that the touch of warm objects promotes social warmth (often also called affection) and prosocial behavior (Williams & Bargh, 2008). In a first experiment, Williams and Bargh (2008) found that participants holding warm beverages perceived a person as socially warmer then when holding cold drinks. In a second experiment, participants were more altruistic, i.e., tended to prefer rewards for other people rather than for themselves, when holding warm vs. cold objects (Williams & Bargh, 2008).

It is assumed that the touch of warmth activates strongly associated concepts of closeness which are learned in early childhood (e.g., a mother's warmth during nursing; see also Harlow, 1958 and Bowlby 1969). Ijzerman and Semin (2009) replicated the findings of Williams and Bargh (2008) and showed that ambient warmth diminishes social distance (i.e., increases social proximity). "When social distance decreases an 'other' is no longer some unknown individual from some anonymous crowd but becomes an identifiable victim" (Thomas C. Schelling, 1968, as cited in Bohnet & Frey, 1999; also see Small & Loewenstein, 2003). As a neurophysiological explanation for their findings, Williams & Bargh (2008) and Ijzerman & Semin (2009) identify the involvement of areas of the insular cortex in both temperature sensations and specific social functions. Kang, Williams, Clark, Gray & Bargh (2010) measured brain activity via fMRI of participants playing the Trust Game while being exposed to warm and neutral thermal conditions. Individuals in the warm condition trusted more than in the neutral condition. Activation patterns of the insular implicated a key function of the insular cortex in mediating the influence of warmth on trust. The insular is strongly associated with empathy, the ability to understand and feel emotions or intentions of others, which is crucial for cooperation and trust (Singer, Seymour, O'Doherty, Kaube, Dolan & Frith, 2004, Jabbi, Swart & Keysers, 2007; Bird, Silani, Brindley, White, Frith & Singer, 2010). This evidence highlights a strong relationship between physical warmth and empathy.

The existing psychological literature forms the research question for our current study: a) Can we replicate the effect of temperature on social behavior using established methods from behavioral economics and b) is it possible to conceptually replicate effects of temperature on social decision making by manipulating ambient room temperature instead of holding cold/warm objects? While the first research question contributes to a larger discussion between psychologists and economists to what extent behavior differs under real vs. hypothetical incentives (for an overview, see Camerer & Mobbs, 2017), the latter is relevant for example in light of the predicted climate change in the next decades. Moreover, it also appears as a more natural way to study temperature effects on behavior.

To address our research questions, we experimentally investigate social decision-making and social perception under different temperature conditions. Indoor thermal sensation primarily depends on ambient temperature, relative humidity and clothing[1] (Parsons, 2019). There is no general agreement about the understanding of thermal perceptive categories like "warm" or "cold". Evidence suggests the optimal performance level at ambient temperatures of 21-22°C (Pilcher, Nadler & Busch's, 2002; Seppanen, Fisk & Lei, 2006; Hancock, Ross, & Szalma, 2007; Yeganeh, Reichard, McCoy, Bulbul, & Jazizadeh, 2018). Ijzerman & Semin (2009) found ambient temperature effects on social distance in 15-18°C and 22-24°C rooms. Kunkel & Kontonasiou (2015) identified for eight European countries that 27°C broadly is considered as warm but still in a potentially comfortable range (see also Givoni, 1992; Yang, Yan & Lam, 2014). Thus, we examined social decision making and social perception in 18°C (as cold), 22°C (as optimal) and 27°C (as warm) environments. Furthermore, we controlled relative humidity and clothing. Humidity affects perceived temperature under extreme temperature conditions. Cold (hot) environments are perceived less comfortable and colder (hotter) when relative humidity is below 30 % (over 50 %; Parsons, 2019). We therefore held relative humidity constant between 30-40 % in all temperature conditions. In a between-subject design manipulating the

[1] Additional relevant factors for thermal reception are thermal radiation (sunshine), wind-chill and skin wetness (sweat). These factors were irrelevant for this experiment, since all measurements were made indoors.

ambient room temperature (cold vs. optimal vs. hot), we measure multiple facets of social behavior within the same subject using methods from both psychology and behavioral economics. Our behavioral economics measures are a Public Good Game, Trust Game, Ultimatum Game, Dictator Game, and a Lying Game (Camerer & Fehr, 2004). Our psychological measures are established concepts related to measurements of Social Warmth, Social Distance and Empathy. Our findings suggest that ambient temperature has no relevant influence on social decision making and social perception. We corroborate our finding of a null effect by the use of equivalence testing.

Hypotheses

Contextual factors can systematically alter behavior for example through framing (Tversky & Kahneman, 1981) or priming (Cohn & Maréchal, 2016) even when cues are subtle and unconscious (Bargh 2002). In our literature review, we found that temperature has a sustained effect on human nature and culture (Wheeler, 1984; Ebensperger, 2001) as well as on human physiology (Parsons, 2014), psychology (Pilcher, Nadler & Busch's, 2002) and behavior (Burke, Hsiang & Miguel, 2015a; 2015b). In particular, we were able to identify three psychological constructs influenced by temperature: First, Williams & Bargh (2008) showed that the touch of warm objects promotes social warmth. Second, Ijzerman & Semin (2009) found similar results altering social distance through ambient warmth. Finally, neuropsychological studies show a strong relationship between temperature sensation and empathy because of shared neural structures (Kang, Williams, Clark, Gray & Bargh, 2010; Singer, Seymour, O'Doherty, Kaube, Dolan & Frith, 2004, Jabbi, Swart & Keysers, 2007; Bird, Silani, Brindley, White, Frith & Singer, 2010). In an attempt to conceptually replicate these effects using an ambient temperature manipulation, we derive the following three hypotheses:

Hypothesis $_{1}^{\textit{Social Warmth}}$***:*** *Increased ambient temperature leads to more social warmth.*

Hypothesis $_{2}^{\textit{Social Distance}}$***:*** *Increased temperature leads to less social distance (i.e., more social proximity).*

Hypothesis $_{3}^{\textit{Empathy}}$***:*** *Increased temperature leads to more empathy.*

Next, we look at the implications on behavior and ask whether an ambient room temperature manipulation is strong enough to influence incentivized decisions. In general, the literature suggests that physical warmth along with social warmth, less social distance and more empathy have positive effects on cooperation and social norm compliance: Williams and Bargh (2008) showed that individuals in a warm condition chose a socially warmer – altruistic – option: Participants briefly holding warm objects more often chose a gift for a friend than for themselves. Similarly, Storey & Workman (2013) found higher cooperation of individuals holding warm objects in an iterated prisoner's dilemma. The literature on interactions between the psychological constructs (social warmth, social distance and empathy) mentioned above and social behavior is even larger (Williams

& Bargh, 2008; Ijzerman & Semin, 2009; Kang, Williams, Clark, Gray & Bargh, 2010; Singer, Seymour, O'Doherty, Kaube, Dolan & Frith, 2004, Jabbi, Swart & Keysers, 2007; Bird, Silani, Brindley, White, Frith & Singer; Dimant, 2019). This motivated our second set of hypotheses which is related to a conceptional replication of the effects of ambient room temperature on behavior across a series of established behavioral economics paradigms related to cooperation and social norm compliance. Our series of behavioral economics tasks included standardized versions of the Dictator Game, Public Goods Game, Trust Game, Ultimatum Game and Lying Game. Since it remains unclear whether changes in behavior in these tasks are driven by changes in preferences and/or changes in beliefs (for example about the motives and intentions of one's partner) we are also eliciting incentivized beliefs, where it is appropriate. This leads to the following set of hypotheses:

Dictator Game

Hypothesis $_{4}^{DG}$**:** *An increase in temperature leads to more altruistic behavior in the Dictator Game.*

Public Goods Game

Hypothesis $_{5}^{PG}{}_{cc}$**:** *Increased temperature leads to more conditional contribution in the Public Goods Game.*

Hypothesis $_{6}^{PG}{}_{beliefs}$**:** *Increased temperature leads to higher beliefs about a partner's contribution.*

Hypothesis $_{7}^{PG}{}_{c}$**:** *Increased temperature leads to higher contributions in the Public Goods Game.*

Trust Game

Hypothesis $_{8}^{TG}{}_{tw}$ *Increased temperature leads to more trustworthiness in the Trust Game.*

Hypothesis $_{9}^{TG}{}_{beliefs}$**:** *Increased temperature leads to higher beliefs about a partner's sent amount.*

Hypothesis $_{10}^{TG}{}_{t}$**:** *Increased temperature leads to more trust in the Trust Game.*

Ultimatum Game

Hypothesis $_{11}^{UG}$: *Decreased temperature leads to less sensitivity to unfair offers in the Ultimatum Game.*

Lying Game

Hypothesis $_{12}^{LG}$: *An increase in temperature leads to less lying.*

Methods

Subjects. We recruited 144 women and 84 men (M_{age} = 22.87 (SD = 4.05), N = 228) from a student subject pool at the University of Mainz using ORSEE; a software platform for organizing and managing experiments (Greiner, 2004). All participants gave written informed consent, participated voluntarily and were completely informed about the procedures of the study. A joint ethics committee formed by Economics & Management Faculties of Universities of Mainz and Frankfurt approved this study.

Procedure. In order to control clothing**,** each subject was asked to change into a t-shirt which was provided upon arrival. Gender-separated changing rooms were available. Participants were then led to their randomly assigned and isolated experimental booths in the climate-controlled computer-laboratory. Room temperature was set to different levels for three experimental groups (M_{cold} = 18.36°C (0.58), $M_{optimal}$ = 22.13°C (0.28), M_{warm} = 27.71°C (0.39)). Humidity was held constant in a comfortable range between 30% and 40% ($M_{humidity|cold}$ = 40.86% (0.53), $M_{humidity|optimal}$ = 36.77% (0.51), $M_{humidity|warm}$ = 32.11% (0.41)). Before testing began, participants stayed in the room for 30 minutes to acclimate to the ambient temperature. During this time participants executed a thirty-minutes Implicit Association Task (IAT) on sustainability (results reported elsewhere). Before beginning the series of economic tasks and social perception measurements, participants were randomly and anonymously paired with a partner. The experimental setup is displayed in Figure 1. At the end of the experiment participants individually left the laboratory room, their body temperature was measured in a neighboring room, and payments were received.

Payment. Participants received a 6 EUR show-up-fee. One task from the series of tasks was randomly chosen for payment. Average earnings were 17.42 EUR for approximately 90 minutes.

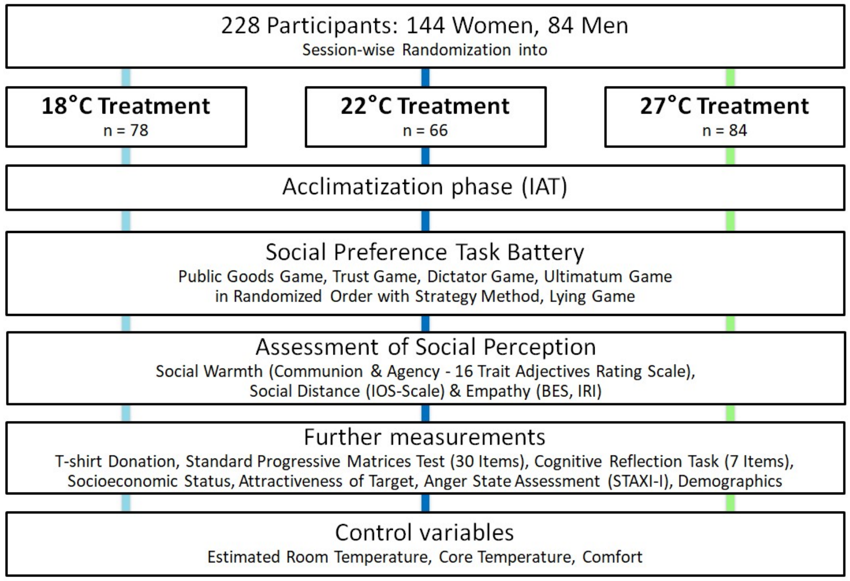

Figure 1 *Experimental Setup*

Measuring social behavior

To measure social preferences, we conducted a series of tasks consisting of a Dictator Game, Ultimatum Game, Public Goods Game, Trust Game and a Lying Game which are well established in the field of behavioral economics (Camerer, 2011).

Public Goods Game. To measure cooperation and reciprocity, we used a Public Goods Game (Marwell, & Ames, 1979). In the Public Goods Game two players are asked to decide on their contribution to a communal project. Participants make two investment decisions. The first contribution decision is made without knowledge about the partner's investment. Second, participants state their contribution as a reaction to each possible investment decisions by the partner (Fischbacher, Gächter & Fehr, 2001). The endowment was 10 EUR, marginal per capita return was 0.75, corresponding to a Prisoner's Dilemma (Kahneman, Knetsch, and Thaler, 1986). **Trust Game.** We measured trust in the partner and trustworthiness using a Trust Game (Berg, Dickhaut & McCabe, 1995). In this game, one subject, the trustor, is asked to send any amount between 0 and 10 EUR to the other subject, the trustee, who receives the tripled amount. The trustee then decides which share of the tripled amount to send back to the trustor. We used the strategy method, so trustees stated their decision as a reaction to each possible decision of the trustor (Charness & Rabin, 2002).

Dictator Game. In order to measure sharing and altruistic behavior a Dictator Game was conducted. In this game participants divide 20 EUR between themselves and a partner (Kahneman, Knetsch & Thaler, 1986). To substantiate the external validity, we observed donation behavior: Participants were given the choice to keep the provided t-shirt or to take it back to their assigned booth, where it would be collected and donated to the German Red Cross.

Ultimatum Game. Furthermore, we used the Ultimatum Game – a take-it or leave-it bargaining game (Güth, Schmittberger & Schwarze, 1982; Camerer & Fehr 2004). One participant, the proposer, offers a share of 20 EUR to a partner, the responder. The responder can then accept the offer or reject it, the latter decision resulting in both participants earning 0 EUR.

Lying Game. To measure honesty and lying behavior we conducted a Lying Game (Fischbacher & Föllmi-Heusi, 2013). In this game participants were instructed to privately roll a die. Subjects were informed that their payoff resulted from the reported roll multiplied by factor 3 when rolling 1 to 5. For rolling a six payout was zero. Cheating could not be identified on individual level in this design.

For measuring the beliefs about other players' behavior, participants were asked to state their assessment of their partners' contribution in the Public Goods Game and the amount sent in the Trust Game, respectively. In order to incentivize correct estimations participants could earn one additional Euro for a correct estimation.

In each task, every participant played both roles and completed each task as a one-shot game. The order of tasks was randomized. For programming the tasks, we used the open-source python-based framework *oTree* (Chen, Schonger, Wickens, 2016).

Measuring social perception

Social warmth. Social warmth is one factor of a universal two-factorial person perception model (Fiske, Cuddy & Glick, 2007). Such dichotomous models have been postulated under a multitude of names and substantially agreed on the same pair of factors (see Abele & Wojciszke, 2014 pp. 199 for an extensive compilation). Abele & Wojciszke (2014) integrated these models into the two factors "communion" (warmth, morality) and "agency" (competence, assertiveness, Wiggins, 1991; Abele & Wojciszke 2007; Hauke & Abele, 2019). The communion and agency dimensions are considered as the Big Two of judgement of self and others in personality psychology (Helm, Abele, Mueller-Kalthoff & Möller, 2017; Abele & Wojciszke, 2018) and were recently considered in economics (Aaker, Vohs & Mogilner 2010; Krings, Sczesny & Kluge, 2011; Jeong, Minson, Yeomans & Gino, 2019). For the assessment of communion and agency in the context of temperature exposure, we used a single item rating scale taken from Abele Uchronski, Suitner & Wojciszke (2008) which consists of 16 trait adjectives. These 16 items were validated in five different countries namely USA, Belgium, Italy, Poland and Germany. The items consist of eight communal and eight agentic adjectives, respectively. Participants made judgements on four different rating targets on a 7-point Likert scale. In addition to rating their assigned partners, participants were asked to rate one female and one male student which were each portrayed on screen. Furthermore, they were asked to imagine and rate a typical student from their university.

Social Distance. To assess social distance (in the sense of interconnectedness) we used Aron, Aron & Smollan's (1992) Inclusion of Others in Self Scale (IOS-Scale, also see: Gächter, Starmer & Tufano, 2015). The IOS-Scale is a one-item language-free measurement (Figure 2). Participants choose from seven pairs of increasingly overlapping circles the one best describing their relationship with their partner, the students in the pictures and the imagined typical student.

Empathy. Since empathy is conceptualized in many ways (Cuff, Brown, Taylor & Howat, 2016), we examined empathy in a multidimensional approach. Recognizing empathy as both a cognitive (ability to understand the emotions of another; Bryant, 1982) as well as an affective (ability to experience the emotions of others; Hogan, 1969) construct, we used the Basic Empathy Scale (BES; Jolliffe & Farrington, 2006; for German version see: Heynen, Van der Helm, Stams & Korebrits, 2016) which measures these two facets. For a further differentiated assessment we used Davis' Interpersonal Reactivity Index (IRI; 1983; for German version see: Paulus, 2009) which distinguishes between four distinctive components: Empathic Concern (feeling emotional concern for others), Perspective Taking (cognitively taking the perspective of another), Fantasy (emotional identification with characters in books, films, etc.), and Personal Distress (negative feelings in response to the distress of others; Pulos, Elison, Lennon, 2004; Davis, 1983).

Control Variables. We conducted measurements of attractiveness of the rating targets. Further we implemented a cognitive ability test, a 7-item cognitive reflection task and assessed participants' socioeconomic status. See the appendix for further information.

Statistical Analysis. The statistical analysis was performed with R in RStudio (2016; R version 3.5.2; RStudio Version 1.1.463).

Manipulation Check. At the end of the experiment participants were asked for their estimates of the room temperature and comfort for controlling the temperature manipulation. Additionally, body temperatures were measured on subjects' foreheads via infrared thermometers immediately after them leaving the computer-laboratory room.

Results

Manipulation Checks

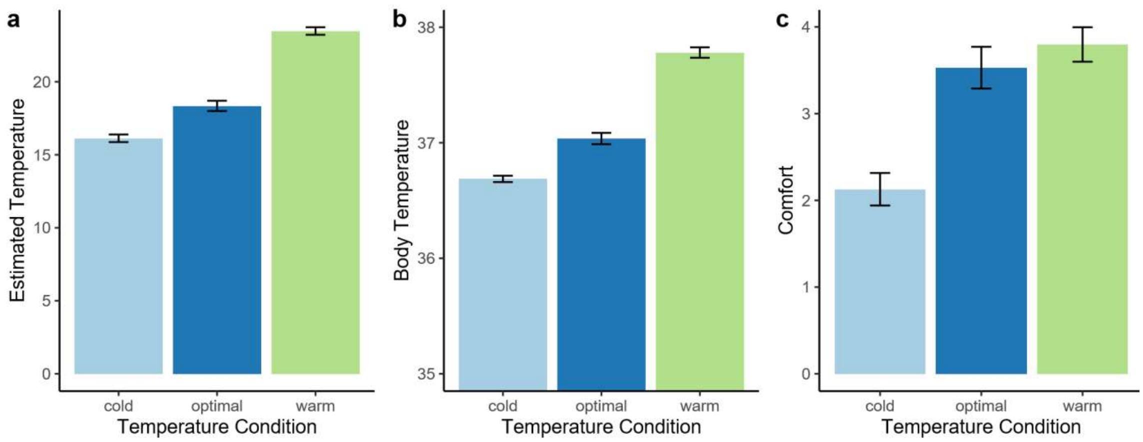

Figure 3 *Estimated temperature, body temperature and comfort ratings by treatment. Error bars represent standard errors of the means.*

In order to check whether the temperature manipulation was successful, we present results of participants' room temperature estimates, body temperature and comfort (Figure 3). Participants' temperature estimates (Figure 3a) were significantly different among all three groups (M_c = 16.13 (2.30), M_o = 18.35 (2.85), M_w = 23.48 (2.37); H =146.69, P < 0.001, n = 228, Kruskal-Wallis Test by Ranks (KWTR)[2]), increasing in temperature and pairwise significantly different ($P_{o\text{-}c}$ < 0.001, $P_{o\text{-}w}$ < 0.001, $P_{c\text{-}w}$ < 0.001, n = 228, Post-hoc Kruskal-Dunn-Test (PKDT)). Body temperature (Figure 3b) differed significantly among temperature conditions (M_c = 36.69 (0.24), M_o = 37.04 (0.40), M_w = 37.78 (0.41); H = 146.37 , P < 0.001, n = 228, KWTR), was increasing in temperature and was pairwise significantly different ($P_{o\text{-}c}$ < 0.001, $P_{o\text{-}w}$ < 0.001, $P_{c\text{-}w}$ < 0.001, n = 228, PKDT).

Self-reported comfort (Figure 3c) differed significantly (M_c = 2.13 (1.65), M_o= 3.53 (1.96), M_w = 3.8 (1.82); H = 38.22 , P < 0.001, n = 228, KWTR). Pairwise comparisons showed significant differences between cold and optimal as well as between cold and warm conditions but not between optimal and warm conditions ($P_{o\text{-}c}$ < 0.001, $P_{o\text{-}w}$ < 0.320, $P_{c\text{-}w}$ < 0.001, n = 228, PKDT). Participants perceived the temperature conditions as different, showed physiological reactions and differed in their comfort judgements. We conclude that the temperature manipulation was successful.

[2] Throughout we report two-tailed test results.

Overall Strategy

Even though the temperature manipulation was successful, we did not find any significant effects of temperature on social behavior and perception. However, absence of evidence does not imply evidence of absence. Minimizing the probability for type 1 errors increases the probability of incorrectly deciding in favor of the null hypothesis (type 2 error) which can be as high as 1 minus the significance level alpha (i.e. 95%; Wellek, 2010; Walker & Nowacki 2011). In order to corroborate the conclusion of a null effect we provide results of nonparametric k-sample tests for equivalence (WWEG, Koh & Cribbie, 2013; Wellek, 2010). Equivalence testing is a well-established tool in fields of medical science and psychology and is increasingly recognized in economics (Kim & Robinson, 2019). The term „equivalence“ in this context means equality except for practically irrelevant differences. Hence, in this sense, a proof of equivalence means that we exclude any medium or high effects (Wellek, 2010)[3]. To test whether regression coefficients are negligible we conduct Two-One-Sided-t-Tests (TOST-β). If the constructed confidence interval around the point estimate is entirely contained in the defined region of practical irrelevance, statistical equivalence can be deduced. However, when testing regression coefficients equivalence bounds cannot be inferred from effect sizes and have to be defined from careful considerations on a practical level to extract the smallest difference of interest (Dixon & Pechmann, 2005; Lakens, Scheel & Isager, 2018). We consider changes of 1% in the outcome variable that are associated with a change of 1°C as negligible. For instance, in a 10-EUR Dictator Game subjects commonly share 3 EUR on average. If an increase in 1°C would results in increase in sharing of 0.03 EUR, we would consider this as practically irrelevant.

[3] Throughout we only report results for $\varepsilon = 0.4$. ε is a bound for the sum of squared non-parametric distances of the k sample distributions.

Social Perception

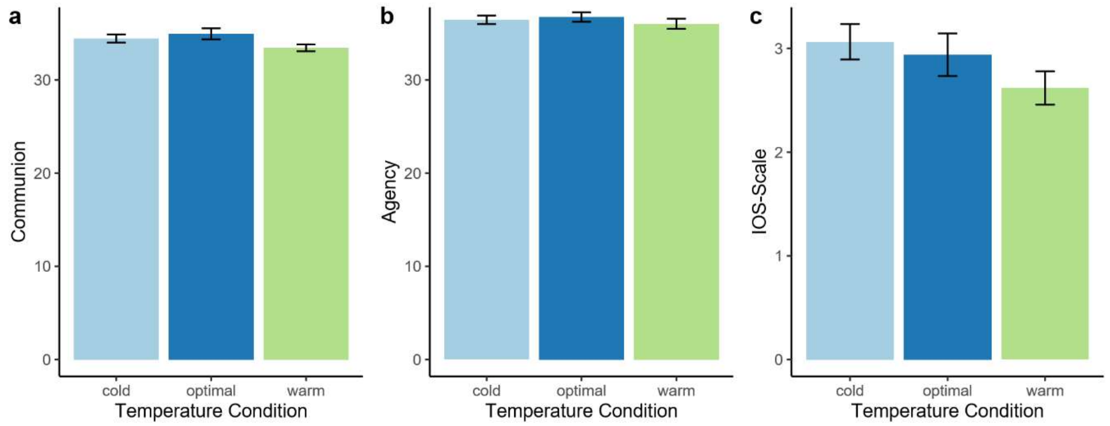

Figure 4 *Average scores in Communion, Agency and IOS-Scale ratings of the partner by treatment. Higher values on the IOS-Scale correspond to a lower reported social distance to the partner. Error bars represent standard errors of the means.*

Communion & Agency. There was no significant effect of temperature on neither communion nor agency scale ratings. This applies to all targets. In the following we report results for participants' ratings of the partner only. Results of the other targets are reported in the appendix. Communion scale ratings of the assigned partner (Figure 4a & 4b) did not differ significantly between groups (M_c = 34.45 (3.82), M_o = 34.95 (4.76), M_w = 33.45 (3.32), H = 3.353, P = 0.187, n = 228, KWTR). Since the rating scale is an aggregation of Likert-scales that only can be interpreted ordinally we used mean ratios for contextual interpretation (Laster & Johnson, 2003). The ratio of means of extreme temperature condition showed a deviation of 2.9% which can be considered as small (Wellek, 2010). Nevertheless, mean comparisons revealed lowest scores in the warm condition which explicitly is against our hypothesis (see medians: Mdn_c = Mdn_o = 33, Mdn_w = 32). An equivalence test was not significant (F = 0.075, P = 0.10; WWEG). The results remain inconclusive; still any potential effect would be contrary to the hypothesis. For all other targets we found significantly equivalent results (see appendix).

We did not observe significant differences between experimental groups in the agency scale ratings (M_c = 36.46 (4.07), M_o = 36.76 (4.08), M_w = 36.04 (4.95); H = 1.177, P = 0.555, n = 228, KWTR). We found identical medians between optimal and warm but not in cold condition (Mdn_c= 36, Mdn_o = Mdn_w = 36.5). Comparison of mean ratios of the temperature extremes revealed a 1.17% deviation, accordingly, the equivalence test was significant (F = 0.013; P = 0.003, WEGG). These results imply no effect of ambient temperature on agency.

Social Distance. We found a tendency in the IOS-Scale ratings for a decrease with temperature (Figure 4c) but the effect was not significant (M_c = 3.06 (1.51), M_o = 2.94 (1.67), M_w = 2.62 (1.47); H =

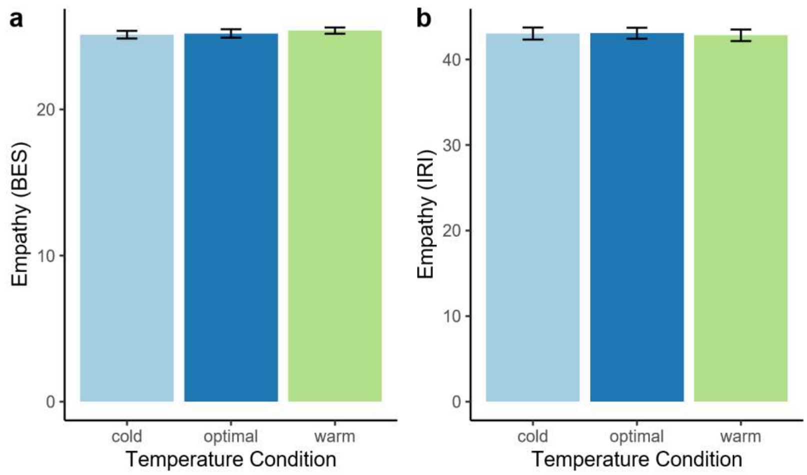

Figure 5 *Average empathy scores in Basic Empathy Scale and Interpersonal Reactivity Index by treatment. Error bars represent standard errors of the means.*

4.066, P = 0.131, n = 228, KWTR). The mean ratio between cold and warm conditions showed a deviation of 14.52%. However, the direction was contrary to our hypothesis and was not significant. Equivalence tests were significant for all other targets (see the appendix for further detail).

Empathy. There were no significant differences in any empathy measure, while all results were significantly equivalent. Figure 5a shows mean Basic Empathy Scale ratings in the three temperature conditions. We found no significant differences between experimental groups (M_c = 25.12 (2.30), M_o = 25.2 (2.36), M_w = 25.39 (1.95); H = 1.07, P = 0.586, n = 228, KWTR), this consequently applies to the subscales: cognitive empathy (M_c = 13.77 (1.45), M_o = 14.14 (1.45), M_w = 13.95 (1.45); H = 2.88, P = 0.237, n = 228, KWTR) and affective empathy (M_c = 11.35 (1.87), M_o = 11.06 (1.76), M_w = 11.44 (1.43); H = 1.57, P = 0.456, n = 228, KWTR). For main and subscales equivalence test results were significant (BES: F = 0.009, P = 0.002; BES cognitive empathy: F = 0.03, P = 0.014; BES affective empathy: F = 0.026, P = 0.010, WWEG).

Further there were no significant effects of temperature on the Interpersonal Reactivity Index (M_c = 43.03 (6.22), M_o = 43.06 (5.21), M_w = 42.82 (6.18); H = 0.074, P = 0.964, n = 228, KWTR; Figure 5b) and none of its subscales (Fantasy Scale (FM): M_c = 14.06 (3.22), M_o = 13.76 (3.12), M_w = 13.75 (3.31); H = 0.579, P = 0.749, n = 228; Empathic Concern (EC): M_c = 14.33 (2.76), M_o = 14.52 (1.97), M_w = 14.36 (2.46); H = 0.051, P = 0.975, n = 228; Perspective Taking (PT): M_c = 14.63 (3.04), M_o = 14.79 (2.51), M_w = 14.71 (2.61); H = 0.029, P = 0.986, n = 228; Personal Distress (PD): M_c = 11.04 (3.07), M_o = 11.91 (2.98), M_w = 11.21 (3.07); H = 2.849, P = 0.241, n = 228, KWTR). The results of equivalence tests for all main and subscales of empathy assessment were significant (Main Scale: F = 0.001, P <

0.001; FS: F_{FS} = 0.006, P_{FS} < 0.001; EC: F_{EC} = 0.003, P_{EC} < 0.001; PT: F_{PT} = 0.002, P_{PT} < 0.001; PD: F_{PD} = 0.042, P_{PD} < 0.03, WWEG).

Social Behavior

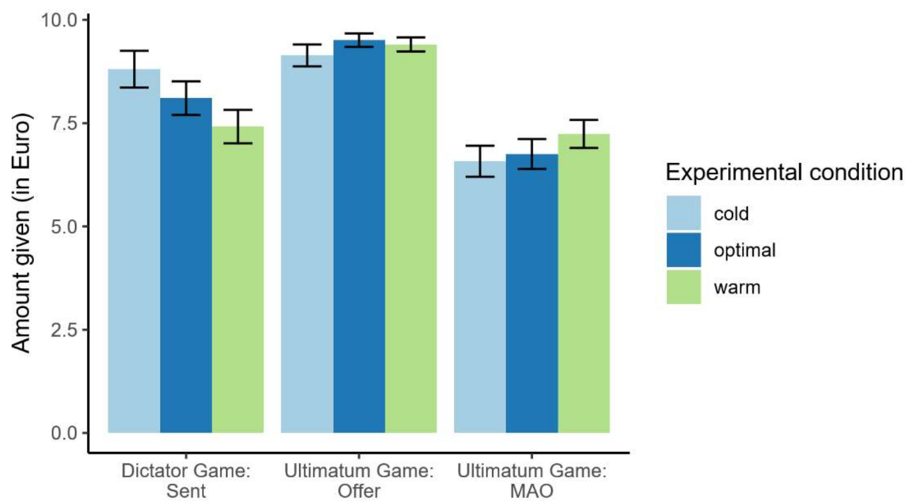

***Figure 6** Mean sent amounts in the Dictator Game, mean offers in the Ultimatum game and mean minimal acceptable offers (MAO) in the Ultimatum Game by treatment. Error bars represent standard errors of the means.*

Dictator Game. Comparing shared amounts in the Dictator Game (Figure 6) showed no significant difference between the three groups (M_c = 8.81 (3.92), M_o = 8.11 (3.30), M_w = 7.42 (3.71); H = 4.02, P = 0.134, n = 228, KWTR). However, mean comparisons revealed a tendency contrary to our hypothesis: Participant give slightly less in warmer conditions. Equivalence testing yielded no significance (F = 0.076, P = 0.11, WWEG) but we found identical medians (Mdn_c = Mdn_o = Mdn_w = 10). Results for sharing measures thus remain statistically inconclusive. Nevertheless, the effect - if existent - would be contrary to the hypothesis that increases in temperature lead to increased social behavior.

Additionally, we observed t-shirt donation decisions in order to assess external validity of altruistic behavior. Participants did not donate significantly more or less between the temperature conditions (M_c = 0.62 (0.49), M_o = 0.67 (0.48), M_w = 0.68 (0.47); H = 0.781, P = 0.677, n = 228, KWTR). Further, medians were significantly equivalent (F = 0.010, P = 0.002, WWEG). Moreover, sharing behavior in the Dictator Game could not predict t-shirt donations (β_{dict} = 0.016, SE = 0.023, P = 0.501, see Table 3 in the appendix).

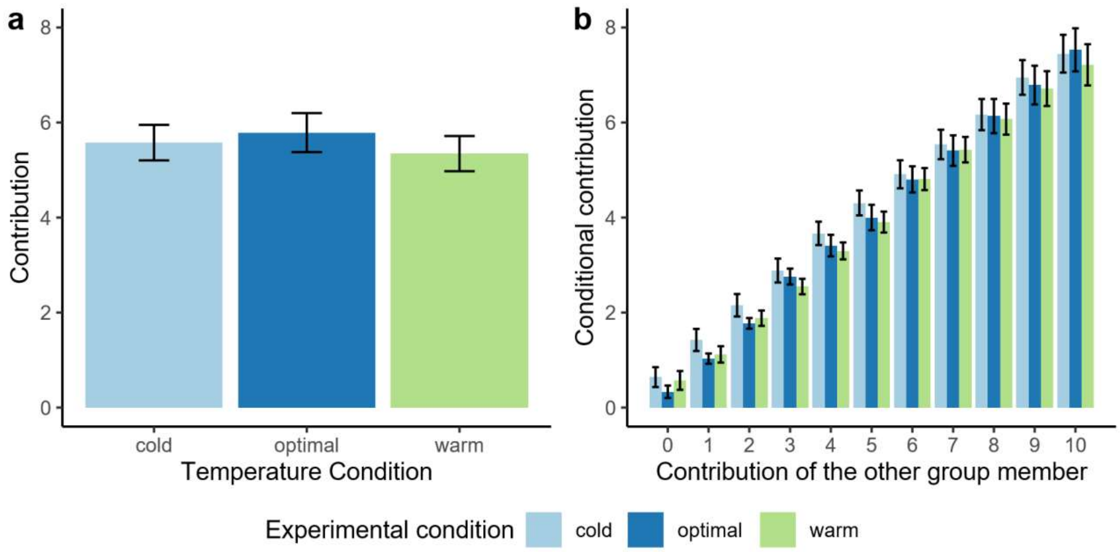

Figure 7 *Mean unconditional and conditional contributions in the Public Good Game by treatment. Error bars represent standard errors of the means.*

Public Goods Game. In order to assess the effect of temperature on cooperation without strategic uncertainty we report results from conditional contributions in the Public Goods Game (Figure 7b.). A Random Effects Regression (RER) of conditional contributions on partner's contribution and temperature controlling for individual effects indicated no significant effect of temperature on conditional contributions in the Public Good Game (RER, β_{temp} = -0.022, SE = 0.03, P = 0.46, Table 1 in the appendix). To test whether the coefficient β_{temp} was negligible a two-one-sided-t-test of β (TOST-β, Dixon & Pechmann, 2005) was conducted. Measurements were made on a scale from 0.00 to 10.00 EUR. In the absence of strategic uncertainty changes of less than one percent for an increase of one degree Celsius are defined as practically irrelevant. Thus, equivalence margins were set accordingly (b_1=0.1, b_2=-0.1). We found β to be significantly equivalent to zero (β = -0.022, SE = 0.03, P = 0.004, TOST-β). A β_{temp} coefficient of -0.022 signifies a decrease of 2.2 cents of mean contribution with a rise in temperature of 1°C, which can be considered as irrelevant. Further, any negative effect would be contrary to our hypothesis. Therefore, we found no differences in cooperation in the Public Good Game based on our a priori formulated hypothesis.

The conditional contributions depend on the beliefs about the partner's unconditional contribution (Kelley & Stahelski, 1970; Falk & Fischbacher, 2006; Fischbacher & Gächter, 2010). We did not find significant differences between groups in the beliefs about the partner's unconditional choice in the Public Good Game (M_c = 5.79 (2.92), M_o = 5.36 (2.94), M_w = 4.96 (3.25); H = 2.787, P = 0.248, n = 228, KWTR). Indeed, we found identical medians of the beliefs in all groups (Mdn_c = Mdn_o = Mdn_w = 5) which were significantly equivalent (F = 0.039, P = 0.025, WWEG).

We found no significant differences in unconditional contributions (Figure 7a) in the Public Good Game between temperature conditions (M_c = 5.58 (3.30), M_o = 5.79 (3.34), M_w = 5.35 (3.40); H = 0.36, P = 0.837, n = 228, KWTR). Indeed, contributions were significantly equivalent (F = 0.009, P = 0.001; WWEG). Additionally, we found identical medians in the temperature conditions (Mdn_c = Mdn_o = Mdn_w = 5) and a difference of means between the extreme temperature conditions ($\Delta_{c\text{-}w}$) of 0.23 EUR. Considering that this difference was found on a scale from 0.00 to 10.00 EUR, the mean difference is only 2.3% for a temperature change of 10°C which we consider economically negligible.

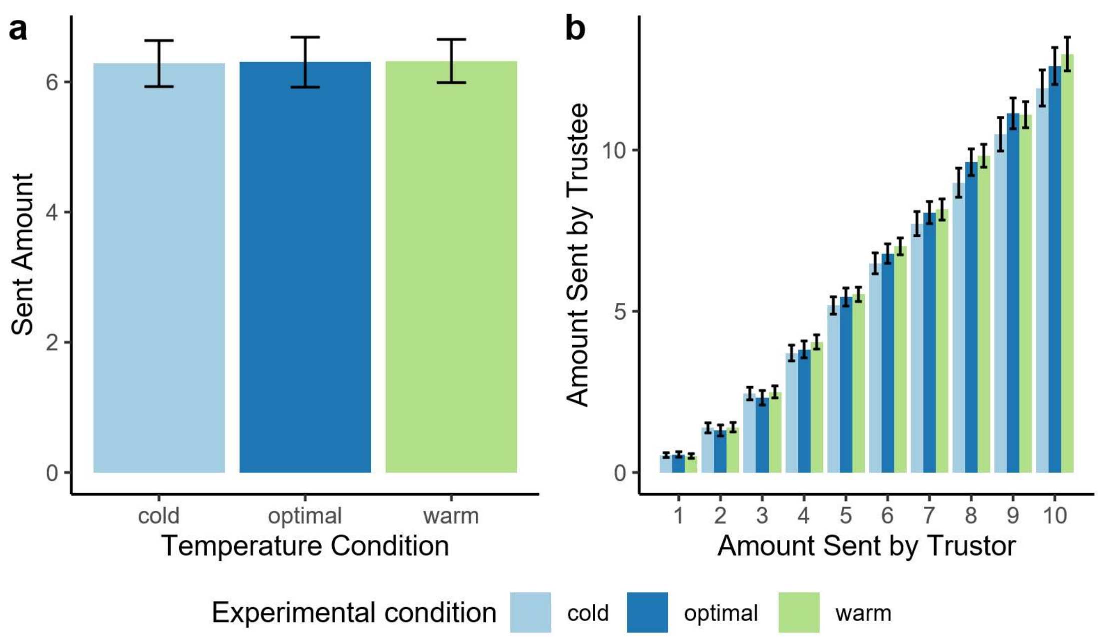

Figure 8 *Mean sent amounts and amounts sent back by trustees in the Trust Game by treatment. Error bars represent standard errors of the mean.*

Trust Game. In order to assess the effect of temperature on trustworthiness without strategic uncertainty we report results on trustworthiness, i.e., the amount sent back conditional on the amount sent by the trustor (Figure 8b.). A RER of trustworthiness on temperature was not significant (β_{temp} = 0.04, SE = 0.035, P = 0.250, see Table 2 in the appendix). We tested for negligibility of β_{temp}. Equivalence margins were set to b_1=0.3 and b_2=-0.3 since measurements were made on a scale from 0.00 to 30.00 EUR (changes less than 1% for a temperature increase of 1°C were considered negligible). The results were significant (β_{temp} = 0.04, SE = 0.035, P < 0.001, TOST-β) and the implied change of 4 cents per one degree Celsius can be considered as practically irrelevant. Therefore, we found no differences in trust and trustworthiness.

Beliefs about the unconditional choice of the partner did not differ significantly between groups (M_c = 8.07 (5.60), M_o = 8.11 (5.25), M_w = 8.33 (6.20); H = 0.011, P = 0.99, n = 228, KWTR) and were significantly equivalent (F = 0.001, P < 0.001, WWEG).

Comparing amounts sent by the trustor between groups in the Trust Game (Figure 8a) we found no significant effect (M_c = 6.28 (3.12), M_o = 6.30 (3.12), M_w = 6.32 (3.04); H = 0.023, P = 0.989, n = 228, KWTR). The equivalence test for sent amounts were highly significant (F = 0.000, P < 0.001; WWEG). We also found identical medians in the optimal and warm condition (Mdn_c = 5, Mdn_o = Mdn_w = 6).

Ultimatum Game. In the Ultimatum Game proposers' offered amounts (Figure 6) were not significantly different (M_c = 9.14 (2.34), M_o = 9.51 (1.31), M_w = 9.40 (1.56); H = 2.11, P = 0.349, n = 228, KWTR). Medians were identical (Mdn_c = Mdn_o = Mdn_w = 10) and the groups were significantly equivalent (F = 0.021, P = 0.007, WWEG). $\Delta_{c\text{-}w}$ was 0.26. Considering that this difference is only 2.6% at a temperature change of 10°C, it can be considered negligible. The MAO of the participants (Figure 6) did not differ significantly between groups (M_c = 6.58 (3.32), M_o = 6.75 (2.93), M_w = 7.24 (3.10); H = 2.081, P = 0.353, n = 228, KWTR). Results of equivalence testing were significant (F = 0.026, P = 0.010, WWEG). $\Delta_{c\text{-}w}$ was 0.66 and medians in the cold and optimal conditions were identical but not in the warm condition (Mdn_c = Mdn_o = 7, Mdn_w = 8). In summary we did not find

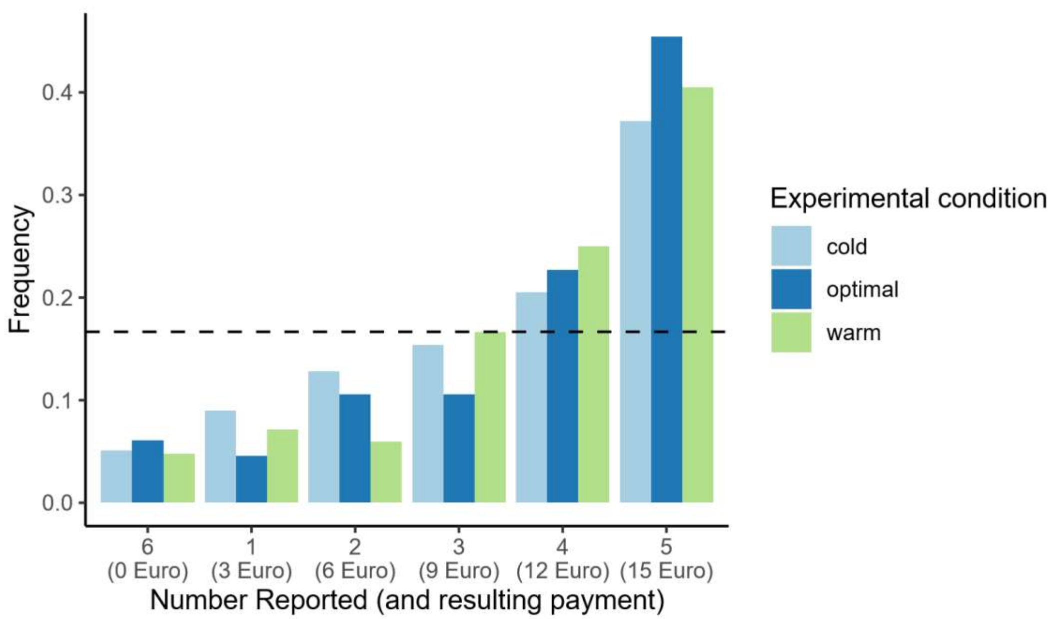

Figure 9 *Frequencies of reported numbers (and resulting payment) in the Lying Game.*

differences in the proposers' offers and minimal acceptable offers.

Lying Game. Figure 9 shows the distributions of reported outcomes in the three experimental conditions. Lying behavior could be observed in all temperature conditions. The proportions of the reported numbers differed significantly from the expected value 16.7% (1/6) in each experimental group (χ^2-goodness-of-fit test, $\chi^2_c = 30.154$, $P_c < 0.000$; $\chi^2_o = 47.455$, $P_o < 0.000$; $\chi^2_w = 49.571$, $P_w < 0.000$). To assess in which group lying was more frequent an Epps-Singleton test (Epps & Singleton, 1986) was conducted. The EST is a distribution test for determining whether two samples have been drawn from the same population. In most cases its power is greater than the Kolmogorov-Smirnov test and it takes the ordinal nature of the data into account (Goerg & Kaiser, 2009). We found no significant differences in lying across temperature conditions (EST, $W_{2\ o\text{-}c} = 2.07$, $P_{o\text{-}c} = 0.722$; $W_{2\ o\text{-}w} = 0.69$, $P_{o\text{-}w} = 0.95$; $W_{2\ c\text{-}w} = 2.03$, $P_{c\text{-}w} = 0.73$).

Gender Effects. We checked for gender effects and found them only for two outcomes: There were differences in comfort ratings between groups and t-shirt donations, see appendix for further detail.

Supplementary Analyses. We found no significant differences in any measurement of social behavior and perception. Indeed, we used equivalence tests to establish practical equivalence of all outcomes. To further corroborate this result, we report additional tests for the absence of the effect in the appendix. These include equivalence tests for different margins and a Bayesian approach based on the Bayes Ratio.

Discussion

In this paper, we test whether ambient warmth influences social perception and behavior. We seek to answer two questions: First, is the positive effect of warm haptic sensations on social warmth, social proximity and empathy transferable to ambient temperature. Second, is this effect strong enough to influence incentivized decision making. Both of these answers have to be answered negatively, although we demonstrated that our temperature manipulation was successful in changing body temperature as well as temperature perception.

We control for relevant factors of thermoreception and use standardized economic methods to rigorously investigate multiple facets of prosocial behavior, specifically cooperation, trust, trustworthiness, altruism, sharing, bargaining and lying. We find that ambient warmth does not affect incentivized decision making or beliefs about one's partners' decisions. Contrary to IJzerman & Semin (2009), we find that ambient warmth does not affect social distance judgements. Also, we find that there is no difference between measurements of communion and empathy across temperature conditions. To corroborate our findings of a null-effect, we used multi group equivalence testing and conclude that all measurements in the Public Goods Game, Trust Game and Ultimatum Game as well as in the social perception measurements are equivalent with respect to an ex-ante specified bound. The results of this study indicate that subtle cueing (Bargh, 2002) via ambient warmth is not strong enough to have a relevant impact on economic behavior.

There are two seemingly contradictory lines of argument for an influence of temperature on human behavior. On one hand, studies indicate that uncomfortably high temperatures impair mood (Keller, Fredrickson, Ybarra, Côté, Johnson, Mikels,... & Wager, 2005), promote aggression (Anderson, 1989; 2001; DeWall, Anderson & Bushman, 2011) and enhance the risk of conflict (Burke, Hsiang & Miguel, 2015b; see also Van Lange, Rinderu & Bushman, 2017). On the other hand, Williams and Bargh (2008) and IJzerman & Semin (2009) provide evidence for a connection between comfortable warmth and prosociality.

Each argument builds upon a well-established theoretical foundation. Heat stress could be considered as a visceral state. Loewenstein (1996) explains changing preferences with hot visceral states like hunger, thirst, sexual desire, moods, emotions, drug cravings, physical pain and fervent emotions. There is ample evidence for a meaningful influence of visceral states on economic behavior. Sleep deprived participants tend to trust less (Anderson & Dickinson, 2010), sexually aroused men are more willing to engage in risky sexual behavior than men who are aren't sexually aroused (Ariely & Loewenstein, 2006) and hungry individuals exhibit lower social preferences (Krause, Ring, Schlichting & Schmidt, 2019). Indeed, in connection with temperature, heat stress caused by a temperature of 30°C leads to more destructive behavior in economic games (Almås, Auffhammer, Bold, Bolliger, Dembo, Hsiang, ... & Pickmans, 2019).

Williams & Bargh (2008) claim that in social interactions, comfortably warm temperatures act as a subtle and unconscious cue that influences thought and behavior (Bargh, 2002). This is often referred to as social priming (Molden, 2014). There are several studies in support of a link between physical warmth and prosociality. For example, there is evidence from neuropsychology showing that temperature perception and social functions like empathy and trust share the same neurological structures (Kang, Williams, Clark, Gray & Bargh, 2010; Singer, Seymour, O'Doherty, Kaube, Dolan & Frith, 2004; Jabbi, Swart & Keysers, 2007; Bird, Silani, Brindley, White, Frith & Singer, 2010). Experimental evidence suggests that subjects with higher body temperatures feel more socially connected to others (Inagaki & Human, 2019). Correlational evidence implies that the physical-to-social warmth link is bidirectional. Zhong & Leonardelli (2008) showed that exclusion could literally feel cold and lonely people tend to self-regulate their feelings through an increased tendency to take warm baths (Bargh & Shalev, 2012). However, social priming studies are under scrutiny (Vadillo, Konstantinidis & Shanks, 2016) and the relationship between temperature and prosociality has been challenged by several failed replication attempts that have been published recently. Neither a high-powered direct replication in three independent locations (Chabris, Heck, Mandart, Benjamin & Simons, 2018) nor a replication in the field (Lynott, Corker, Wortman, Connell, Donnellan, Lucas & O'Brien, 2014) could find a promoting effect of warm objects on prosocial behavior. Matching their field data with weather data, Lynott, Corker, Connell, & O'Brien (2017) aimed to control for the influence of ambient temperature and were not able to establish conclusive results.

So far, all studies on the effect of temperature on prosocial behavior relied on costless or hypothetical behavior. We provide an experimental investigation of prosociality as a trade-off between egoistic concerns and concerns for a partner using monetary incentives. Incentivization reduces possible "presentation effects" (i.e., generosity; Camerer & Hogarth, 1999) which is crucial in order to gauge the practical relevancy of an effect. Any effect on social behavior that is of practical relevance to economists should have a common cause and be present when the stakes are real. As exposure to the ambient temperature is ubiquitous and varying haptic sensations are less common, we provide a conceptual replication of the physical-to-social warmth link via ambient temperature.

The replication crisis in the field of social sciences has been acknowledged and publications of replication studies have increased exponentially in the economic literature over the last years (OSC, 2015; Camerer, Dreber, Forsell, Ho, Huber, Johannesson & Heikensten, 2016). The field of social priming has been called "the poster child for doubts about the integrity of psychological research" (Kahneman, 2012). Notoriously, social priming effects are subtle and highly sensible to experimental context as well as the subject pool (Cesario, 2014; Yu, Abrams & Zacks, 2014). Yet, Bargh (2014) insists that "unconscious influences [of social priming] on judgement, emotion, behavior and motivation are of practical importance both to society as a whole and to the everyday lives of its members". Our findings represent a strong contradict of this claim, at least when it comes to temperature effects. We showed that several facets of social behavior and perception are statistically equivalent across temperature conditions.

If a temperature effect of adequate size were to exist but our design were inadequate to measure it because priming effects occur only under unique circumstances, then these situations would probably be too rare to be considered practically relevant. There are many known artefacts confounding experimental research like experimenter demand effects (Zizzo, 2010), demographic effects (Casari, Ham, & Kagel, 2007), incentive size effects (Slonim & Roth, 1998), framing (Kahneman & Tversky, 1981) and priming effects (Cohn & Maréchal, 2016). Our results show that ambient temperature does not represent a confounding effect. This implies that to the extent that temperature is in a range between 17°C and 28°C a temperature-controlled environment is not necessary for incentivized social behavior experiments and experiments should be comparable across different temperature regions.

Naturally, one replication is just “one data point among many that bear on this question” (Bargh & Melnikoff, 2019) and our results do not disprove that temperature may have a subtle effect on certain variables. However, the results from this study and from other replications strongly challenge the existence of the physical-to-social warmth connection and indicate the need for further elaboration and adjustment of the theoretic model.

In the social sciences there is a strong focus on significance rather than relevance (Ziliak & McCloskey, 2008; Duvendach, Palmer-Jones, Reed, 2017). It is essential for economic research not only to prove statistical significance but to put effects and effect sizes into context. The paradigm of solely relying on null-hypothesis testing is increasingly criticized (Kim & Ji, 2015; Rao & Lovric, 2016), not least because the p-value is a decreasing function of the sample-size (Kim & Robinson, 2019; Ohlson, 2019). For large enough samples, any miniscule effect will become significant. Equivalence testing (i.e., interval-based hypothesis testing) may be a good alternative which is already a well-established tool in fields of medical science and psychology. Equivalence testing is increasingly recognized in economic science (Kim & Robinson, 2019) and holds the potential to determine whether an observed effect is statistically negligible and can be economically ignored (Lakens, Scheel & Isager, 2018). However, equivalence margins are always context dependent and should be carefully determined from theory and meaningful effect sizes should be considered (Lakens, Scheel & Isager, 2018; Kim & Robinson, 2019). There is a grey space between negligibility and meaningfulness, in which effects can exist. Through equivalence testing we are able to determine whether such an effect is relevant or can be economically ignored.

Appendix

A Additional results

Communion and Agency

In the following we present the results of the communion and agency measurements for the picturized female student, the picturized male student and the imagined typical student.

Female Student. There were no significant effects of temperature on communal ratings of the female student (M_c = 41.62 (5.90); M_o = 41.56 (5.63); M_w = 41.05 (5.07), H = 0.572, P = 0.751, n = 228, KWTR) and further no significant effects on agency ratings (M_c = 39.01 (5.16); M_o = 39.26 (5.17); M_w = 39.48 (4.46), H = 0.234, P = 0.890, n = 228, KWTR). We found significant equivalence between groups for communion (F = 0.007, P = 0.001; WWEG) and agency (F = 0.005, P < 0.001; WWEG).

Male Student. Differences of communion ratings of the male student were not significant (M_c = 38.5 (5.06); M_o = 38.26 (5.02); M_w = 37.24 (3.89), H = 2.570, P = 0.277, n = 228, KWTR), however they were significantly equivalent (F = 0.044, P = 0.032; WWEG). Agency ratings did not differ significantly, as well (M_c = 38.92 (4.33); M_o = 38.88 (4.21); M_w = 38.48 (4.46), H = 0.467, P = 0.792, n = 228, KWTR), then again equivalence testing showed significant results (F = 0.007, P = 0.001; WWEG).

Typical Student. We found no significant effects in communion scale ratings of the imagined typical student (M_c = 36.24 (4.56); M_o = 36.7 (5.44); M_w = 35.70 (4.4), H = 1.556, P = 0.459, n = 228, KWTR). Results of the agency scale ratings also were not significant (M_c = 37.36 (4.79); M_o = 38.68 (4.37); M_w = 38.32 (4.68), H = 2.752, P = 0.253, n = 228, KWTR). Equivalence tests for agency (F = 0.043, P = 0.030; WWEG) and communion (F = 0.021, P = 0.007) were significant.

IOS-Scale. Neither for the imagined student target we found significant effects (Mc = 4.09 (1.43); Mo = 3.79 (1.34); Mw = 3.68 (1.38), H = 3.954, P = 0.139, n = 228, KWTR) nor for the pictured male student (Mc = 3.10 (1.49); Mo = 2.83 (1.34); Mw = 2.8 (1.36), H = 1.753, P = 0.416, n = 228, KWTR) and the pictured female student (Mc = 3.28 (1.48); Mo = 3.33 (1.42); Mw = 3.12 (1.35), H =1.012, P = 0.603, n = 228, KWTR). Equivalence tests of IOS-Scale ratings for all other targets were significant (typical student: F = 0.049, P = 0.041; WWEG; pictured male student: F = 0.029, P = 0.013; WWEG; pictured female student F = 0.047, P = 0.037; WWEG).

Ability. Raven's Standard Progressive Matrices. We conducted a 30-items variation of the Raven's Standard Progressive Matrices Test (SPM; Raven & Court, 1958) to assess cognitive ability. In order to further increase the difficulty of the tasks, four items of the original SPM were randomly chosen from section A, five from section B, six from section C, seven from section D and eight from section E. The participants had 10 minutes to work through the matrices tasks. Ability of participants did not differ significantly between groups (M_c = 24.42 (3.30); M_o = 24.48 (2.79); M_w = 24.83 (2.54), H = 0.627, P = 0.731, n = 228, KWTR), however results were significant in the equivalence test (F = 0.054, P = 0.003; WWEG).

Overconfidence. Neither overestimation (the estimation of own ability versus real own ability) nor overplacement (estimation of own ability versus estimation of others' ability) differed significantly between temperature conditions (overestimation: (M_c = -0.42 (2.96); M_o = -1.41 (4.6); M_w = -0.64 (3.20), H = 1.024, P = 0.599, n = 228, KWTR; overplacement: M_c = 1.92 (2.59); M_o = 1.62 (3.43); M_w = 1.57 (2.5), H = 0.81, P = 0.67, n = 228, KWTR). Equivalence tests for overestimation and overplacement were significant (overestimation: F = 0.038, P = 0.024, WWEG; overplacement: F = 0.009, P = 0.002; WWEG).

CRT. We conducted a 7 items Cognitive Reflection Task (Toplak, West & Stanovich, 2011; 2014) and a 10 minute variation of the Raven's Standard Progressive Matrices Test with 30 items to assess cognitive ability. Participants were asked for their estimations of their own and the average performance of the other participants in the SPM for overestimation measurement. We found no significant differences in CRT results (M_c = 4.18 (1.99); M_o = 3.89 (2.06); M_w = 3.98 (1.97), H = 0.823, P = 0.663, n = 228, KWTR), however results of the equivalence test were significant (F = 0.010, P = 0.002; WWEG).

STAXI. In order to control anger we used the anger-state part of the State-Trait-Anger Expression Inventory (Spielberger et al., 1983). By 10 items anger states were assessed on a 4-point Likertscale. Results for Anger State measurement were not significant (M_c = 1.68 (2.14); M_o = 1.64 (3.40); M_w = 1.00 (1.45), H = 3.720, P = 0.156, n = 228, KWTR), equivalence test, however, was significant (F = 0.054, P = 0.049; WWEG).

SES. We assessed socioeconomic status with three items in the childhood (Griskevicius, Delton, Robertson & Tybur, 2011). The assessment of socioeconomic status showed no significant differences (M_c = 10.94 (4.45); M_o = 10.39 (3.68); M_w = 10.15 (4.32), H = 3.720, P = 0.156, n = 228, KWTR). The means were significantly equivalent (F = 0.054, P = 0.006; WWEG).

Attractiveness. We measured perceived attractiveness to rule out any effects on the rating targets. We found no significant differences of attractiveness between experimental conditions and ratings of any target (partner: M_c = 4.09 (0.40); M_o = 4.03 (0.55); M_w = 3.93 (0.58), H = 5.355, P = 0.069, n =

228, KWTR; female student: $M_c = 4.35$ (1.19); $M_o = 4.50$ (1.11); $M_w = 4.74$ (1.13), $H = 4.937$, $P = 0.085$, $n = 228$, KWTR; male student: $M_c = 3.85$ (1.23); $M_o = 3.91$ (1.11); $M_w = 3.93$ (1.26), $H = 0.11$, $P = 0.944$, $n = 228$, KWTR; typical student: $M_c = 4.32$ (1.05); $M_o = 4.48$ (0.93); $M_w = 4.56$ (1.18), $H = 1.68$, $P = 0.43$, $n = 228$, KWTR). However, we found significant results in equivalence tests of attractiveness on the assigned partner ($F = 0.053$, $P = 0.048$, WWEG), the picturised male student ($F = 0.003$, $P < 0.001$, WWEG) and the typical student ($F = 0.028$, $P = 0.012$, WWEG) but not for the female student ($F = 0.063$, $P = 0.070$, WWEG).

B Robustness checks of the equivalence margin ε

Economic Test Series

	Variable	F	$P_{\varepsilon=0.3}$	$P_{\varepsilon=0.4}$	$P_{\varepsilon=0.5}$
Dictator Game	Sharing	0.076	0.341	0.106	0.018
Prosocial Behaviour	T-Shirt Donation	0.010	0.019	0.002	0.000
Ultimatum Game	Offer	0.021	0.052	0.007	0.001
	MAO	0.026	0.068	0.010	0.001
Public Goods Game	Contribution	0.009	0.015	0.001	0.000
	Beliefs	0.039	0.130	0.025	0.003
Trust Game	Trust	0.000	0.000	0.000	0.000
	Beliefs	0.001	0.002	0.000	0.000

Social Perception Measures

	Variable	F	$P_{\varepsilon=0.3}$	$P_{\varepsilon=0.4}$	$P_{\varepsilon=0.5}$
Communion	Partner	0.075	0.334	0.103	0.017
	Male picturised student	0.044	0.154	0.032	0.004
	Female picturised student	0.007	0.011	0.001	0.000
	Typical student	0.021	0.052	0.007	0.001
Agency	Partner	0.013	0.027	0.003	0.000
	Male picturised student	0.007	0.011	0.001	0.000
	Female picturised student	0.005	0.007	0.001	0.000
	Typical student	0.043	0.146	0.030	0.003
Social Distance	Partner	0.047	0.170	0.047	0.004
	Male picturised student	0.029	0.082	0.013	0.001
	Female picturised student	0.013	0.025	0.003	0.000
	Typical student	0.049	0.181	0.041	0.005
Empathy	BES	0.009	0.016	0.002	0.000
	BES – cognitive	0.030	0.088	0.014	0.001
	BES – affective	0.026	0.069	0.010	0.001
	IRI	0.001	0.001	0.000	0.000
	IRI – Fantasy Scale	0.006	0.010	0.001	0.000
	IRI – Empathetic Concern	0.003	0.004	0.000	0.000
	IRI – Perspective Taking	0.002	0.002	0.000	0.000
	IRI – Personal Distress	0.042	0.145	0.030	0.003

C Bayes Factors

In order to corroborate the conclusion of a null effect we further provide Bayes factors. Bayesian analysis has become increasingly popular in social sciences and holds, among other advantages, the potential to receive evidence in favor of a null hypothesis (Kass and Raftery, 1995). We calculated Bayes factors for all main-effects for ANOVA designs (Rouder, Morey, Speckman & Province, 2012). A Bayes factor greater than 1 can be considered as strong evidence for the alternative hypothesis, a Bayes factor close to 1 remains inconclusive while Bayes factors close to 0 can be considered as strong evidence for the null-hypothesis (Jeffreys, 1939). In the latter case, Jeffreys (1939) specifies that Bayes factors between 1/3 and 1/10 indicate moderate evidence for the null-hypthesis and Bayes factors less than 1/10 can be considered as strong evidence for the null-hypothesis (van Doorn, van den Bergh, Bohm, Dablander, Derks, Draws, ... & Ly, 2019).

Economic Game

	Variable	**Bayes Factor**	**Evidence for H_0**
Dictator Game	Sharing	0.163	moderate
	T-Shirt Donation	0.050	strong
Ultimatum Game	Offer	0.093	strong
	MAO	0.141	moderate
Public Good Game	Contribution	0.075	strong
	Conditional Contributions	0.002	strong
	Beliefs	0.333	moderate
Trust Game	Trusting	0.058	strong
	Trustworthiness	0.004	strong
	Beliefs	0.076	strong

Social Perception

	Variable	**Bayes Factor**	**Evidence for H_0**
Communion	Partner	0.224	moderate
	Male picturised student	0.066	strong
	Female picturised student	0.053	strong
	Typical student	0.079	strong
Agency	Partner	0.070	strong
	Male picturised student	0.052	strong
	Female picturised student	0.068	strong
	Typical student	0.282	moderate
Social Distance	Partner	0.115	moderate
	Male picturised student	0.090	strong
	Female picturised student	0.059	strong
	Typical student	0.065	Strong

	Variable	Bayes Factor	Evidence for H_0
Empathy	BES	0.057	strong
	BES – cognitive	0.147	moderate
	BES – affective	0.108	moderate
	IRI	0.055	strong
	IRI – Fantasy Scale	0.051	strong
	IRI – Empathic Concern	0.063	strong
	IRI – Perspective Taking	0.062	strong
	IRI – Personal Destress	0.192	moderate

Control variables

	Variable	Bayes Factor	Evidence for H_0
Attractiveness	Partner	0.092	strong
	Male picturised student	0.064	strong
	Female picturised student	0.667	inconclusive
	Typical student	0.095	strong
Anger	STAXI-I	0.134	moderate
Ability	Raven's SPM	0.410	inconclusive
	CRT	0.081	strong
Social Economic Status	SES	0.121	moderate

D Tables

Table 1: Random Effects Regression of the Conditional Contribution in the Public Goods Game.

	Dependent variable:
	Cooperation
Contribution	0.692***
	(0.010)
Temperature	-0.022
	(0.029)
Constant	1.060
	(0.657)
Observations	2,508
R^2	0.655
Adjusted R^2	0.654
F Statistic	4,749.629***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: Random Effects Regression of Trustworthiness in the Trust Game.

	Dependent variable:
	Trustworthiness
Trust	1.354***
	(0.014)
Temperature	0.040
	(0.035)
Constant	-2.209***
	(0.785)
Observations	2,280
R^2	0.809
Adjusted R^2	0.808
F Statistic	9,613.559***

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Probit Regression of Shirt Donation Decisions on Giving in the Dictator Game.

	Dependent variable:	
	Shirt Donation	
	(1)	(2)
Male		0.465**
		(0.187)
Dictator Giving	0.015	0.026
	(0.023)	(0.024)
Constant	0.270	0.020
	(0.204)	(0.233)
Observations	228	228
Log Likelihood	-146.885	-143.687
Akaike Inf. Crit.	297.771	293.373

Note: *p<0.1; **p<0.05; ***p<0.01

Literature

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.

Abele, A., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology, 38 (7)*, 1202–1217.

Abele, A., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93 (5)*, 751-763.

Abele, A., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in Experimental Social Psychology, 50,* 195–255.

Abele, A., & Wojciszke, B. (2018). *Agency and communion in social psychology.* Routledge.

Almås, I., Auffhammer, M., Bold, T., Bolliger, I., Dembo, A., Hsiang, S. M., . . . Pickmans, R. (2019). *Destructive behavior, judgment, and economic decision-making under thermal stress.* National Bureau of Economic Research.

Anderson, C., & Dickinson, D. L. (2010). Bargaining and trust: The effects of 36-h total sleep deprivation on socially interactive decisions. *Journal of Sleep Research, 19 (1)*, 54–63.

Anderson, C. A. (1989). Temperature and aggression: Ubiquitous effects of heat on occurrence of human violence. *Psychological Bulletin, 106 (1)*, 74-96.

Anderson, C. A. (2001). Heat and violence. *Current Directions in Psychological Science, 10 (1)*, 33–38.

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making, 19 (2)*, 87–98.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social psychology, 63 (4)*, 596-612.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41 (3)*, 258-290.

Bargh, J. A. (2002). Losing consciousness: Automatic influences on consumer judgment, behavior, and motivation. *Journal of Consumer Research, 29 (2)*, 280–285.

Bargh, J. A. (2014). Our unconscious mind. *Scientific American, 310 (1)*, 30–37.

Bargh, J. A., & Melnikoff, D. (2019). Does physical warmth prime social warmth? Reply to Chabris et al. (2019). *Social Psychology, 50 (3)*, 207–210.

Bargh, J. A., & Shalev, I. (2012). The substitutability of physical and social warmth in daily life. *Emotion, 12 (1)*, 154-162.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10 (1)*, 122–142.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133 (5)*, 1515–1525.

Bohnet, I., & Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review, 89 (1)*, 335–339.

Bouchama, A., & Knochel, J. P. (2002). Heat stroke. *New England Journal of Medicine, 346 (25)*, 1978–1988.

Bowlby, J. (1969). Attachment and loss: Volume i: Attachment. In *Attachment and loss: Volume i: Attachment* (pp. 1–401). London: The Hogarth Press and the Institute of Psycho-Analysis.

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, 413–425.

Büntgen, U., Tegel, W., Nicolussi, K., McCormick, M., Frank, D., Trouet, V., . . . Wanner, H. et al. (2011). 2500 years of European climate variability and human susceptibility. *Science, 331 (6017)*, 578–582.

Burke, M., Hsiang, S. M., & Miguel, E. (2015a). Climate and conflict. *Annual Review of Economics, 7 (1)*, 577–617.

Burke, M., Hsiang, S. M., & Miguel, E. (2015b). Global non-linear effect of temperature on economic production. *Nature, 527 (7577)*, 235-239.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., . . . Chan, T. et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351 (6280)*, 1433–1436.

Camerer, C. F., & Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies, 97*, 55–95.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*, 7–42.

Camerer, C., & Mobbs, D. (2017). Differences in behavior and brain activity during hypothetical and real choices. *Trends in cognitive sciences*, *21 (1)*, 46-56.

Cao, M., & Wei, J. (2005). Stock market returns: A note on temperature anomaly. *Journal of Banking & Finance, 29 (6)*, 1559–1573.

Casari, M., Ham, J. C., & Kagel, J. H. (2007). Selection bias, demographic effects, and ability effects in common value auction experiments. *American Economic Review, 97 (4)*, 1278–1304.

Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science, 9* (1), 40–48.

Chabris, C., Heck, P. R., Mandart, J., Benjamin, D. J., & Simons, D. J. (2018). No evidence that experiencing physical warmth promotes interpersonal warmth: Two failures to replicate Williams and Bargh (2008). Social Psychology, 50 *(2)*, 127-132.

Charkoudian, N. (2003). Skin blood flow in adult human thermoregulation: How it works, when it does not, and why. *Mayo Clinic Proceedings, 78 (5)*, 603–612.

Charness, G., & Gneezy, U. (2008). What's in a name? anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization, 68 (1)*, 29–35.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics, 117 (3)*, 817–869.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9,* 88–97.

Cohn, A., & Maréchal, M. A. (2016). Priming in economics. *Current Opinion in Psychology, 12,* 17–21.

Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review, 8 (2)*, 144–153.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44 (1)*, 113-126.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics, 4 (3)*, 66–95.

DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2011). The general aggression model: Theoretical extensions to violence. *Psychology of Violence, 1 (3)*, 245-258.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, *73*, 66-88.

Dixon, P. M., & Pechmann, J. H. (2005). A statistical test to show negligible trend. *Ecology*, *86 (7)*, 1751-1756.

Duvendack, M., Palmer-Jones, R., & Reed, W. R. (2017). What is meant by "replication" and why does it encounter resistance in economics? *American Economic Review, 107 (5)*, 46–51.

Ebensperger, L. A. (2001). A review of the evolutionary causes of rodent group-living. *Acta Theriologica, 46 (2)*, 115–144.

Epps, T., & Singleton, K. J. (1986). An omnibus test for the two-sample problem using the empirical characteristic function. *Journal of Statistical Computation and Simulation, 26 (3)*, 177–203.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54 (2)*, 293–315.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association, 11 (3)*, 525–547.

Fischbacher, U., & Gachter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review, 100 (1)*, 541–56.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71 (3)*, 397–404.

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.

Fountoulakis, K. N., Savopoulos, C., Zannis, P., Apostolopoulou, M., Fountoukidis, I., Kakaletsis, N., . . . Tsikerdekis, A. et al. (2016). Climate change but not unemployment explains the changing suicidality in Thessaloniki Greece (2000–2012). *Journal of Affective Disorders, 193*, 331–338.

Gächter, S., Starmer, C., & Tufano, F. (2015). Measuring the closeness of relationships: A comprehensive evaluation of the 'inclusion of the other in the self' scale. *PloS ONE, 10 (6)*, e0129478.

Givoni, B. (1992). Comfort, climate analysis and building design guidelines. *Energy and Buildings, 18 (1)*, 11–23.

Goldman, R. F. (2001). Introduction to heat-related problems in military operations. *Medical Aspects of Harsh Environments, 1*, 3–49.

Greiner, B. (2004). An Online Recruitment System for Economic Experiments. In: Kurt Kremer, Volker Macho (Eds.): *Forschung und wissenschaftliches Rechnen 2003*. GWDG Bericht 63, Göttingen: Ges. für Wiss. Datenverarbeitung, 79-93, 2004.

Griskevicius, V., Delton, A. W., Robertson, T. E., & Tybur, J. M. (2011). Environmental contingency in life history strategies: The influence of mortality and socioeconomic status on reproductive timing. *Journal of Personality and Social Psychology, 100 (2)*, 241-254.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3 (4)*, 367–388.

Hancock, P. A., Ross, J. M., & Szalma, J. L. (2007). A meta-analysis of performance response under thermal stressors. *Human Factors, 49 (5)*, 851–877.

Harlow, H. F. (1958). The nature of love. *American Psychologist, 13 (12)*, 673-685.

Hauke, N., & Abele, A. (2019). Two faces of the self: Actor-self perspective and observer-self perspective are differentially related to agency versus communion. *Self and Identity, 19 (1),* 1–23.

Holopainen, J., Helama, S., Björkenstam, C., & Partonen, T. (2013). Variation and seasonal patterns of suicide mortality in Finland and Sweden since the 1750s. *Environmental Health and Preventive Medicine*, *18 (6)*, 494-501.

Helm, F., Abele, A., Mueller-Kalthoff, H., & Möller, J. (2017). Applying dimensional comparison theory to the fundamental dimensions of social judgment – agency and communion. *Learning and Individual Differences, 54*, 116–125.

Heyes, A., & Saberian, S. (2019). Temperature and decisions: Evidence from 207,000 court cases. *American Economic Journal: Applied Economics, 11 (2)*, 238–265.

Heynen, E., Van der Helm, G., Stams, G., & Korebrits, A. (2016). Measuring empathy in a German youth prison: A validation of the German version of the Basic Empathy Scale (BES) in a sample of incarcerated juvenile offenders. *Journal of Forensic Psychology Practice, 16 (5)*, 336–346.

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., . . . Engelbrecht, F. et al. (2018). Chapter 3: Impacts of 1.5°C global warming on natural and human systems. In *Intergovernmental Panel on Climate Change (IPCC) Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* World Meteorological Organization: Geneva, Switzerland, 2018.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33 (3)*, 307-316.

Holopainen, J., Helama, S., & Partonen, T. (2014). Does diurnal temperature range influence seasonal suicide mortality? Assessment of daily data of the Helsinki metropolitan area from 1973 to 2010. *International Journal of Biometeorology, 58 (6)*, 1039–1045.

IJzerman, H., & Semin, G. R. (2009). The thermometer of social relations: Mapping social proximity on temperature. *Psychological Science, 20 (10)*, 1214–1220.

Inagaki, T. K., & Human, L. J. (2019). Physical and social warmth: Warmer daily body temperature is associated with greater feelings of social connection. *Emotion*. Advance online publication. https://doi.org/10.1037/emo0000618

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage, 34 (4)*, 1744–1753.

Jeffreys, H. (1939). *The theory of probability*. OUP Oxford.

Jeong, M., Minson, J., Yeomans, M., & Gino, F. (2019). Communicating with warmth in distributive negotiations is surprisingly counterproductive. *Management Science, 65 (12),* 5813-5837.

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence, 29 (4)*, 589–611.

Kahneman, D. (2012). A proposal to deal with questions about priming effects. Retrieved November 29, 2019. Retrieved from https://www.nature.com/news/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20etter.pdf

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review, 76 (4)*, 728–741.

Kahneman, D., & Tversky, A. (1981). On the study of statistical intuitions. *Cognition, 12,* 325–326.

Kang, Y., Williams, L. E., Clark, M. S., Gray, J. R., & Bargh, J. A. (2010). Physical temperature effects on trust behavior: The role of insula. *Social Cognitive and Affective Neuroscience, 6 (4)*, 507–515.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90 (430)*, 773–795.

Keller, M. C., Fredrickson, B. L., Ybarra, O., Côté, S., Johnson, K., Mikels, J., . . . Wager, T. (2005). A warm heart and a clear head: The contingent effects of weather on mood and cognition. *Psychological Science, 16 (9)*, 724–731.

Kelley, H. H., & Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology, 16 (1)*, 66-91.

Kim, J. H., & Ji, P. I. (2015). Significance testing in empirical finance: A critical review and assessment. *Journal of Empirical Finance, 34 (1)*, 1–14.

Kim, J. H., & Robinson, A. P. (2019). Interval-based hypothesis testing and its applications to economics and finance. *Econometrics, 7 (2)*, 21-43.

Koh, A., & Cribbie, R. (2013). Robust tests of equivalence for k independent groups. *British Journal of Mathematical and Statistical Psychology*, *66 (3)*, 426-434.

Krause, S., Ring, P., Schlichting, K. & Schmidt, U. (2019). *First food, then morals - the impact of food deprivation on social behavior*. Unpublished.

Krings, F., Sczesny, S., & Kluge, A. (2011). Stereotypical inferences as mediators of age discrimination: The role of competence and warmth. *British Journal of Management, 22 (2)*, 187–201.

Kunkel, S., & Kontonasiou, E. (2015). Indoor air quality, thermal comfort and daylight policies on the way to nZEB—Status of selected MS and future policy recommendations. *Proceedings of the ECEEE Summer Study, First Fuel Now, Belambra Les Criques, Toulon, France*, 1–6.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1 (2)*, 259-269.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4 (2)*, 195–208.

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychological Bulletin, 136 (6)*, 1045-1067.

Laster, L. L., & Johnson, M. F. (2003). Non-inferiority trials: The 'at least as good as' criterion. *Statistics in Medicine, 22 (2)*, 187–200.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65 (3)*, 272–292.

Lynott, D., Corker, K. S., Connell, L., & O'Brien, K. S. (2017). The effect of haptic and ambient temperature experience on prosocial behavior. *Archives of Scientific Psychology, 5 (1)*, 10-18.

Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. B., Lucas, R. E., & O'Brien, K. (2014). Replication of "Experiencing physical warmth promotes interpersonal warmth" by Williams and Bargh (2008). *Social Psychology, 45,* 216-222.

Marwell, G., & Ames, R. E. (1979). Experiments on the provision of public goods. i. resources, interest, group size, and the free-rider problem. *American Journal of Sociology, 84 (6)*, 1335–1360.

Molden, D. C. (2014). Understanding priming effects in social psychology: What is "social priming" and how does it occur? *Social Cognition, 32*, 1–11.

Ohlson, J. A. (2019). Researchers' data analysis choices: An excess of false positives? *Available at SSRN 3221894*.

Okamoto-Mizuno, K., Mizuno, K., Michie, S., Maeda, A., & Iizuka, S. (1999). Effects of humid heat exposure on human sleep stages and body temperature. *Sleep, 22 (6)*, 767–773.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349 (6251)*, 943-950.

Ormandy, D., & Ezratty, V. (2012). Health and thermal comfort: From who guidance to housing strategies. *Energy Policy, 49*, 116–121.

Parsons, K. (2014). *Human thermal environments: The effects of hot, moderate, and cold environments on human health, comfort, and performance*. CRC Press.

Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index.

Pilcher, J. J., Nadler, E., & Busch, C. (2002). Effects of hot and cold temperature exposure on performance: A meta-analytic review. *Ergonomics, 45 (10)*, 682–698.

Pulos, S., Elison, J., & Lennon, R. (2004). The hierarchical structure of the Interpersonal Reactivity Index. *Social Behavior and Personality, 32 (4)*, 355–360.

Rao, C. R., & Lovric, M. M. (2016). Testing point null hypothesis of a normal mean and the truth: 21st century perspective. *Journal of Modern Applied Statistical Methods, 15 (2)*, 3-21.

Raven, J. C. & Court, J. (1958). *Standard Progressive Matrices: Sets A, B, C, D & E*. HK Lewis London.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56 (5)*, 356–374.

Schelling, T. C. (1968). The life you save may be your own. *Problems in Public Expenditure*, 127–162.

Seppanen, O., Fisk, W. J., & Lei, Q. (2006). *Effect of temperature on task performance in office environment*. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA (US).

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303 (5661)*, 1157–1162.

Slonim, R., Roth, A. E. et al. (1998). Learning in high stakes ultimatum games: An experiment in the Slovak republic. *Econometrica, 66*, 569–596.

Small, D. A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty, 26 (1)*, 5–16.

Spielberger, C. D., Gorsuch, R., Lusene, R. E., Vagg, P. R., & Jacobs G. A. (1983). *Manual for the State Trait Anxiety Inventory.* Consulting Psychologists Press; Palo Alto, CA (US).

Storey, S., & Workman, L. (2013). The effects of temperature priming on cooperation in the iterated prisoner's dilemma. *Evolutionary Psychology, 11 (1)*, 52–67.

Team, R. (2016). Rstudio: Integrated development for R [computer software]. *URL http://www.rstudio.com/. Boston, MA: RStudio, Inc.*

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20 (2)*, 147–168.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39 (7)*, 1275-1289.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211 (4481)*, 453–458.

Vadillo, M. A., Konstantinidis, E., & Shanks, D. R. (2016). Underpowered samples, false negatives, and unconscious learning. *Psychonomic Bulletin & Review, 23 (1)*, 87–102.

Van Lange, P. A., Rinderu, M. I., & Bushman, B. J. (2017). Aggression and violence around the world: A model of climate, aggression, and self-control in humans (clash). *Behavioral and Brain Sciences, 40 (1),* 1-63.

van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., ... & Ly, A. (2019). The JASP guidelines for conducting and reporting a Bayesian analysis.

Wang, X. (2017). An empirical study of the impacts of ambient temperature on risk taking. *Psychology*, *8 (7)*, 1053-1062.

Walker, E., & Nowacki, A. S. (2010). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine, 26* (2), 192–206.

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. Chapman and Hall/CRC.

Wheeler, P. E. (1984). The evolution of bipedality and loss of functional body hair in hominids. *Journal of Human Evolution, 13 (1)*, 91–98.

Wiggins, J. S. (1991). *Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior*. University of Minnesota Press.

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science, 322 (5901)*, 606–607.

Yang, L., Yan, H., & Lam, J. C. (2014). Thermal comfort and building energy consumption implications – a review. *Applied Energy, 115*, 164–173.

Yeganeh, A. J., Reichard, G., McCoy, A. P., Bulbul, T., & Jazizadeh, F. (2018). Correlation of ambient air temperature and cognitive performance: A systematic review and meta-analysis. *Building and Environment, 143*, 701–716.

Yu, A. B., Abrams, R. A., & Zacks, J. M. (2014). Limits on action priming by pictures of objects. *Journal of Experimental Psychology: Human Perception and Performance, 40* (5), 1861-1873.

Zhong, C.-B., & Leonardelli, G. J. (2008). Cold and lonely: Does social exclusion literally feel cold? *Psychological Science, 19 (9)*, 838–842.

Ziliak, S., & McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. University of Michigan Press.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13 (1)*, 75–98.