The effects of temperature on prosocial and antisocial behaviour: A review and meta-analysis

Dermot Lynott [1, 3], Katherine Corker [2], Louise Connell [1 3] and Kerry O'Brien [4].

[1] Department of Psychology, Maynooth University

[2] Department of Psychology, Grand Valley State University

[3] Department of Psychology, Lancaster University

[4] School of Social Sciences, Monash University

Author Note

We have no conflicts of interest to declare. Correspondence concerning this article should be addressed to dermot.lynott@mu.ie

A big thank you to the following people for providing additional information on studies or raw/summary data for the meta-analyses: David la Porte, Francesca Citron, Rebecca Parker, Hans IJzerman, Adam Fay, Rachel Callicoat, Blythe Duell, Lawrence Williams, Craig Anderson, Anna Steidle, Christine Gockel.

Word count: 10,382

Keywords: temperature, prosocial, antisocial, behaviour, priming, aggression

Data availability statement: The data that support the findings of this study are openly available in the project's OSF page: https://osf.io/xpyzj/?view_only=8b3288219cbb487397a0fdd1c3260d23

NOTE: This is a preprint version of an article that has been accepted for publication at the British Journal of Social Psychology

Abstract

Research from the social sciences suggests an association between higher temperatures and increases in antisocial behaviours, including aggressive, violent, or sabotaging behaviours, and represents a *heat-facilitates-aggression* perspective. More recently, studies have shown that higher temperature experiences may also be linked to increases in prosocial behaviours, such as altruistic, sharing, or cooperative behaviours, representing a *warmth-primes-prosociality* view. However, across both literatures, there have been inconsistent findings and failures to replicate key theoretical predictions, leaving the status of temperature-behaviour links unclear. Here we review the literature and conduct meta-analyses of available empirical studies that have either prosocial (e.g., monetary reward, gift giving, helping behaviour) or antisocial (self-rewarding, retaliation, sabotaging behaviour) behavioural outcome variables, with temperature as an independent variable. In an omnibus multivariate analysis (total N = 4577) with 80 effect sizes, we found that there was no reliable effect of temperature on the behavioural outcomes measured. Further, we find little support for either the *warmth-primes-prosociality* view or the *heat-facilitates-aggression* view. There were no reliable effects if we consider separately the type of behavioural outcome (prosocial or antisocial), different types of temperature experience (haptic or ambient), or potential interactions with the experimental social context (positive, neutral or negative). We discuss how these findings affect the status of existing theoretical perspectives, and provide specific suggestions advancing research in this area.

Keywords: temperature, prosocial, antisocial, behaviour, priming, aggression

The effects of temperature on prosocial and antisocial behaviour: A review and meta-analysis

Temperature is an inescapable feature of the environment. Global temperatures are rising, and although the consequences of this rise for the physical environment have long been apparent (NASA, 2017; Peterson et al., 2009), behavioural consequences are now also being highlighted. The role of temperature in influencing behaviour is particularly important due to the multitude of both individual and population-level effects that have been demonstrated (e.g., Anderson, 2001; Hsiang, et al., 2013; Huang et al., 2014; IJzerman et al., 2013; Kang, et al., 2011; Williams & Bargh, 2008). If temperature variation does lead to robust and predictable effects on human behaviour, it is critical to understand properly the mechanisms underpinning temperature's relationship to social interactions and behaviour.

Currently, two broad, but competing, perspectives exist concerning how temperature influences social interactions and behaviour, each with contrasting predictions and their own evidence base. Some experimental and field-based research has tended to associate higher temperature experience with increases in aggression, violence, and non-aggressive antisocial responding (e.g., Anderson & Anderson, 1984, 1986; Anderson, et al., 2000; Anderson et al., 1997; Fay & Maner, 2014; Kenrick & McFarlane, 1986; Vrij, et al., 1994). Such patterns have been supported by epidemiological, population-based research, which has observed associations between higher temperature experience and increases in antisocial behaviour, violence, and societal volatility (e.g., Burke et al, 2009; Hsiang et al., 2011, 2013; Kelley et al., 2015; Larrick et al., 2011; Zhang et al., 2007). On the other hand, experimental social psychological research has had a greater tendency to find and report associations of higher temperatures with greater prosociality, social connectedness, trust, and altruism (e.g., Bargh & Shalev, 2012; Huang et al., 2014; IJzerman, et al., 2013; IJzerman & Semin, 2009, Kang et al., 2011; Miyajima & Meng, 2017; Storey & Workman, 2013; Williams & Bargh, 2008).

Thus, we have two literatures, which have generally kept to themselves, each with their own way of doing things, and each of which has seen failures to replicate key findings (e.g., Buhaug, 2010; Donnellan et al., 2015; Lynott et al., 2014; McCarthy, 2014; Wortman et al., 2014). However, if a stimulus to which people are constantly exposed, such as temperature has the potential to influence a range of behavioural outcomes, then the impact of the (ir)reproducibility of its effect goes far beyond lab and discipline-specific debates and becomes essential to our understanding of what is true about human behaviour.

The goal of this paper is to investigate if temperature truly affects behaviour, and we investigate three critical questions: Overall, how large are the effects of temperature on social behaviour reported in the literature? Are higher temperatures associated with increases in prosocial behaviour, antisocial behaviour, or both? Does the magnitude of reported effects depend on moderators such as type of behaviour being measured, the form of temperature manipulation, or social context of the study? To answer these questions, we review the available evidence and conduct a meta-analysis of existing empirical research. Whereas some meta-analyses in psychology have been hampered by the effects of publication bias (e.g., overestimating effect sizes) by focusing purely on published peer-reviewed work (Carter & McCullough, 2014 ; Van Elk et al., 2015), we overcome such methodological barriers by incorporating both published and unpublished work, and by using multiple statistical techniques to estimate and counteract possible biases in the sample of studies selected, in an attempt to provide convergent evidence for the presence, or absence, of temperature effects. We also report separate meta-analyses for prosocial and antisocial behavioural outcomes. This series of meta-analyses will therefore provide a best estimate of the likely effect size of the impact of temperature on both prosocial and antisocial outcomes. As such, the findings will quantify the evidence for current conflicting theoretical frameworks and allow us to conclude which framework (if any) can explain how temperature affects social behaviour.

Temperature-Behaviour Frameworks

In the following sections we review supporting empirical evidence, and counterevidence, for the effects of temperature, first in terms of the *heat-facilitates-aggression view*, and then in terms of the *warmth-primes-prosociality* view.

Heat-Facilitates-Aggression View

Perhaps the most prominent psychological theory relating to the *Heat-Facilitates-Aggression View* is the *heat hypothesis* developed by Anderson and colleagues (Anderson, 2001; Anderson et al., 1997). As van Lang, Rinderu and Bushman (2017) summarise: temperature is “a factor that triggers aggression and violence (General Aggression Model)”. Stated simply, this view suggests that higher temperatures lead people to feel increased discomfort and negative affect, which then leads to increased arousal. However, people misattribute their discomfort and arousal to those around them, who then become a target for their aggressive motives, thoughts, and behaviours (see Anderson et al., 1995, 1996; Anderson & Bushman, 2002). The general relationship espoused by the heat-hypothesis is supported by much empirical evidence, from laboratory-based experimental studies to population-based longitudinal and cross-sectional studies.

In the 1970s, a series of studies by Baron and Bell (Baron, 1972, 1976; Baron & Bell, 1975, 1976; Bell & Baron, 1976, 1977) showed that higher temperatures resulted in people making more aggressive responses to a partner in experimental settings. For example, in hotter conditions, people would direct longer and more intense noise blasts at another participant (Baron & Bell, 1975). It is worth noting that findings from some of these studies suggested that if where the environmental or social context was more negative or highly aversive (referred to as "negative affect") at higher temperatures (e.g., if participants were confronted with an angry confederate), participant aggression would decrease, as participant motives changed from a "fight" to a "flight" response in order to remove themselves from the situation and reduce their overall negative affect (e.g., Baron, 1972). While many of these early studies did not explicitly measure escape motives, Palmarek and Rule (1979) did just that, providing partial evidence in support of the *negative affect escape* model. On the one hand, they found that insulted participants avoided an aggressive task at higher temperatures (i.e., they wanted to escape), but with another measure being more equivocal about participants' motives, where there was no difference in participants' rated preference for the aggressive vs non-aggressive task. Moreover, since these initial studies, findings in the literature have more consistently supported the general heat-facilitates-aggression view (e.g., van Goozen, Frijda, Kindt, & van de Poll, 1994; Vrij & van der Steen, 1994).

Indeed, several studies have exposed this more general trend, by extending the type of behaviors influenced by temperature beyond aggression to incorporate irritable or antisocial behaviors more broadly construed. Kenrick and MacFarlane (1986) found a linear relationship between temperature and driver irritability, such that higher temperatures were associated with an increase in the frequency of horn honking when drivers were obstructed at a four-way stop sign. In two studies looking at aggression in baseball (Reifman, et al., 1991; Larrick et al., 2011), higher temperatures were associated with more aggressive play, and increased instances of retaliation, where pitchers would directly target batters. In a laboratory simulation study, Vrij and colleagues found that police officers exhibited more aggressive shooting behaviours (e.g., having a gun in their hand, rather than in their holster) in a hot compared to a cold environment (Vrij et al., 1994). In a lab study examining anger-provoking situations and the intensity of aggressive responding, people were more likely to withhold money from a researcher at higher temperatures (van Goozen, et al., 1994), indicating more negative evaluations from participants. Indeed, the amount of money given was strongly negatively related to the anger levels displayed by participants, such that higher levels of anger were associated with greater reduction in the sums of money being offered. More recently, Fay and Maner (2014) found that experience of higher temperatures (sitting on a heated seat pad) lead to increased feelings of hostility – participants produced a higher number of aggressive words in a word-stem completion task and directed more intense noise blasts at a study partner in hot relative to neutral conditions.

Some additional support for a link between higher temperatures and aggression comes from epidemiological research that considers effects of climate change over wide geographic areas and over much longer timescales than are normally considered in psychological research. Anderson and colleagues observed associations between higher temperatures and increases in violent crime across many datasets (Anderson, 1989; Anderson & Anderson, 1984, 1996), using different time frames, geographic locations, and various dependent measures. For example, state capitals in the U.S. with higher annual temperatures exhibited higher rates of violent crime compared to states with lower mean temperatures (Anderson, 1987; Anderson et al., 2000). Supporting this pattern, Ranson (2014, p. 274) also showed that temperature has a "strong positive effect on criminal behaviour", estimating that in the coming decades, increases in global temperature are likely to cause approximately 11.7 million additional instances of crime in the US alone. Similar patterns have been observed by many other groups looking at conflict over wide geographical areas, including effects of temperature on conflict in Syria (Kelley et al., 2015), civil wars in Africa (Burke et al., 2009) and across the globe (Hsiang et al., 2011). A meta-analysis collating evidence from 60 studies covering a period of ~10,000 years showed that for each standard deviation increase in temperatures, the rates of interpersonal and intergroup violence were associated with 4% and 14% increases, respectively (Hsiang et al., 2013).

Some researchers have suggested that findings supporting a *heat-facilitates-aggression view*, may be explained by alternative theoretical proposals. For example, the *routine activities view* (Cohen & Felson, 1979) suggests that increases in violence and crime may in fact be driven by changes in people's everyday activities in summer months compared to winter months. For example, in warmer months, people are more likely to be outside of their homes, leading to more opportunities for violence or criminal activity. Although changes in daily activities may be an explanatory factor in some instances (e.g., Field, 1992), they cannot explain the increased aggression observed in baseball games played at higher temperatures (Larrick et al., 2011), the increased shooting activity of police officers in hot conditions (Vrij et al., 1994), the fact that violent crime increases during hotter days within summer months (Anderson & Anderson, 1984), or indeed any lab-based observations of temperature-aggression links, where routine activities have no role to play.

Thus, many studies, both experimental and epidemiological, can be construed as supporting a general *heat-facilitates-aggression* perspective.

Counter-Evidence to the Heat-Facilitates-Aggression View

Despite the quantity of evidence that supports a link between higher temperatures and increased aggression, conflict, and antisocial behaviour more generally, the larger picture is not so clear-cut. Many of the paradigms mentioned above, have produced counterevidence or null results elsewhere. For example, contrary to Kenrick and McFarlane (1986), Baron (1976) found that levels of horn-honking were not predicted by temperature, but depended on the behaviour of a confederate pedestrian crossing the road (e.g., whether they were on crutches or provocatively dressed). In contrast to Bell and Baron (1975), Baron (1972) found that participants were more likely to apply shocks to a partner in a cool condition than participants in a hot condition, regardless of the experimental context.

In terms of population-based studies, some have found that conflict actually increases during colder periods. For example, increases in conflict (over the period 1400-1900) were associated with cooler periods in the earth's history (Zhang et al., 2007), a finding supported by independent analysis (Tol & Wagner, 2010). Additionally, it has been suggested that claimed associations between higher temperatures and increases in violence and societal instability (e.g., Burke et al., 2009) are partly due to statistical anomalies and partly due to poor operational definitions of basic concepts such as conflict, drought, and even heat (Buhaug, 2010). Rather, more appropriate statistical analyses and better definitions of conflict terms reveal no relationship between temperature and conflict, with political and economic factors being far stronger predictors of violence, conflict, and war (Buhaug, et al., 2014).

Lastly, a meta-analysis of lab-based studies conducted prior to 1995 found that the overall effect of temperature on aggression was not reliably different from zero (Anderson et al., 2000). However, laboratory effects were moderated by differences in social context, with neutral contexts showing a weak, but statistically significant effect (Cohen's $d = 0.26$), whereas positive and negative contexts (e.g., where the experimenter or confederate behaved positively or negatively towards participants) did not reveal reliable effects.

Thus, while the effects of heat on aggressive behaviour have become received wisdom, to the extent that there are many references in introductory psychology textbooks (e.g., Hewestone et al., 1996; Krahé, 2013; Pakes, 2009; Plotnik, 2013), support for the *heat-facilitates-aggression* view is far from universal.

Warmth-Primes-Prosociality View

A very different approach, *social priming*, has been proposed as a mechanism for how higher temperature experiences lead to more prosocial behaviour. In what has become a modern classic of experimental social psychology, having been cited over 1,800 times, Williams and Bargh (2008) observed that participants who briefly interacted with a warm, as opposed to a cold object (e.g., hot coffee vs iced coffee), were more likely to behave prosocially towards a target individual or a friend. Williams and Bargh suggest that positive developmental experiences of warmth (e.g., through bonding between caregiver and infant) intrinsically associate physical and interpersonal warmth. According to this view, when we experience physical warmth we automatically and unconsciously activate concepts associated with interpersonal warmth, and are therefore predisposed to act more prosocially. The findings of Williams and Bargh align with the view that “warmth” is seen as a core dimension on which we evaluate other people (Fiske et al, 2007; Koch, et al., 2020; cf. Nauts et al., 2014) and provides a basis for explaining a range of social judgements and interactions. Many studies have since built on this perspective, and provided evidence for associations between higher temperature, positive affect, and prosocial behaviours (IJzerman et al., 2013; Kang et al., 2011; Storey & Workman, 2012) or indeed for links between greater interpersonal closeness and perceptions of higher environmental temperatures (Bargh & Shalev, 2012; Leander et al., 2012; Szymkow et al., 2013).

Following Williams and Bargh, Kang and colleagues (Kang et al., 2011) found that participants handling a warm therapeutic pack (compared to a cold one) showed greater trust towards an anonymous “partner”, whose role was played by a computer program. Using a similar manipulation, Storey and Workman (2013) found that when people handled a warm object compared to a cold one, they were more likely to cooperate in an iterated prisoners’ dilemma paradigm. Such patterns have also been extended to the developmental domain where IJzerman and colleagues (2013) found that children who scored highly on a secure attachment scale behaved more prosocially (offering gifts for other children) in warmer compared to colder conditions. By contrast, there was no effect of the temperature manipulation for children who were low in secure attachment.

Beyond social priming, IJzerman and colleagues (e.g., Beckes et al., 2015; Hu et al., 2019; IJzerman et al., 2015) have outlined a social thermoregulation view, which suggests that because the regulation of body temperature is critical for survival more generally, it may be that sensitivity to temperature cues is involved in regulating social interactions as well as physical. For example, IJzerman et al. (2012) found that following a socially excluding incident, participants experienced a decrease in skin temperature. Furthermore, compensating for a decrease in skin temperature (e.g., by holding a warm cup of tea) counteracted the negative affect that accompanies social exclusion. Certain observations, such as the positive relationship between the size of social networks and core body temperatures, suggest that thermoregulation may be fundamental to many aspects of social development (IJzerman et al., 2015).

Thus, whether one assumes a social priming or thermoregulation mechanism underlying temperature-prosocial behaviour links, the above studies seem to provide a consistent narrative that links higher temperature experiences, either via ambient or transient haptic experience, to increased prosocial responding.

Counterevidence for the Warmth-Primes-Prosociality View

Despite the myriad examples that higher temperature experiences may increase prosocial outcomes, the evidence from the literature reveals this relationship to be weak and inconsistent across comparable studies. In replicating both studies from Williams and Bargh (2008), Chabris and colleagues (Chabris et al., 2019) found no evidence that experiencing physical warmth promoted interpersonal warmth or prosocial behaviour. Furthermore, three high-powered replications of Williams and Bargh's second study (2008), could not replicate the effect that a brief interaction with a warm object led to more prosocial responding than an interaction with a cold object (Lynott et al., 2014). Subsequent Bayesian analysis of these effects showed better support for the null model (i.e., no effect of pack temperature) compared to the alternative model (Lynott et al., 2017; Dienes & McLatchie, 2018). However, in an analysis of the effects of *ambient* temperature on the same prosocial outcome measure (choosing a gift for a friend), a significant association between higher temperatures and greater prosocial responding was observed for one participant sample, but no effect of ambient temperature for a second sample (Lynott et al., 2017; see also IJzerman et al., 2013).

As well as null effects, sometimes increased temperatures are associated with reduced prosocial responding. In a paradigm examining employee interactions with customers, employees showed increased prosocial behaviour towards consumers (by giving larger discounts) at *lower* temperatures (Kolb et al., 2012), a pattern replicated elsewhere (e.g., Belkin & Kouchaki, 2017). Such a pattern is not consistent with a *warmth-primes-prosociality* perspective, although others have argued that colder temperatures trigger a greater need for affiliative behaviours (e.g., Van Acker et al., 2016), which subsequently leads to greater prosociality, or that greater fatigue at higher temperatures may reduce positive affect, which then leads to a reduction in prosociality (Belkin & Kouchaki, 2017). Overall however, these patterns suggest that the *warmth-primes-prosociality* framework may not be as robust as some have assumed.

Methodological Shortcomings and Reproducibility

In terms of methodological issues, there are some commonalities between the *warmth-primes-prosociality* and the *heat-facilitates-aggression* perspectives, including generally small sample sizes for lab-based studies. Although sample sizes in experimental social psychology have been increasing over the past decade (e.g., Schilder et al.,; 2014; Lynott et al; 2014), many studies have historically relied on low numbers of 20-50 participants per condition, using between-participant designs to investigate whether higher temperatures cause either aggressive/antisocial behaviour (e.g., Bell 1980; Boyanowsky, 1999; Fay and Maner, 2014; Gockel et al., 2014) or prosocial behaviour (e.g., Kolb, Gockel & Werth, 2012; Leander et al., Chartrand & Bargh, 2012; Storey & Workman, 2013; Williams & Bargh, 2008). Small samples mean that such studies are quite underpowered for detecting even moderate effect sizes, and that any observed effect sizes will have wide confidence intervals surrounding estimates. What's more, small samples are more likely to lead to a higher false discovery rate (Button et al., 2013) and, where *p*-values for effects are hovering around the critical $p = .05$ significance level, the results may provide more evidence for a null effect than for the alternative hypothesis (Lakens & Evers, 2014). However, the development of an array of statistical tools (e.g., PET-PEESE, *p*-curve, trim and fill) allows researchers to determine whether these patterns reflect genuine effects or if they are likely to be the result of some other forces (e.g., selective reporting, *p*-

hacking). The utility of such tools has been demonstrated (Lakens et al., 2016; Peters et al., 2007; Stanley & Doucouliagous, 2014), with some being successfully applied to recent meta-analyses, for example in the domain of religious priming studies (Shariff et al., 2015). We make use of these tools in the current meta-analysis.

As well issues stemming from small samples, there are also broader investigatory blind spots to consider within each framework. The broad goal of the *heat-facilitates-aggression* research enterprise has been to identify links between temperature and aggression or negative affect. However, given that the behavioural effects of temperature are potentially much broader than just finding a locus in aggressive responding, it is surprising that this work has not sought to address the other side of this coin and include measurements of non-aggressive and prosocial responding. Similarly, the many empirical studies (largely since the mid 2000s), that have attempted to find links between higher temperatures and more prosocial and positive affective outcomes, have focused almost exclusively on prosocial outcomes, without considering antisocial responding in the same contexts. Thus, there has been a rather complete siloisation of investigations of prosocial and antisocial responding, which sets empirical, and therefore theoretical limits on the interpretation of these findings. That is, can a theory that predicts an increase in aggressive or antisocial behaviour at higher temperatures also predict an accompanying decrease in prosocial behaviour? Or vice versa? To our knowledge, such predictions have not been specified or tested. To gain a better understanding of the possible breadth of temperature on behaviour, we must take a broader view, and attempt to integrate currently disjoint perspectives.

A final limitation is that work at the population level has focused solely on negative outcomes (societal volatility, antisocial behaviour, violence, conflict. E.g., Burke et al., 2009, Hsiang et al., 2013; Ranson, 2014), without considering whether there may also be comparable effects on positive aspects of human behaviour. Anderson et al., (1995) raised a similar point with respect to experimental work, bemoaning the narrowness of the methodological focus of temperature-behaviour studies and emphasizing the need for a broader perspective, which has failed to clearly emerge in the two decades intervening.

Current Meta-Analysis

The focus of the current meta-analysis is on lab-based experiments and experimental field studies, and extends previous meta-analyses in the temperature-behaviour domain in several ways. First, we address the broad question of how higher temperatures affect both prosocial and aggressive/antisocial behaviours, in contrast to previous meta-analyses that examined only aggressive outcomes (Anderson et al., 2000; Hsiang et al., 2013), or measures such as person perception, moral judgements and self-regulation that do not relate directly to pro- or anti-social behaviour (IJzerman et al., 2021). Second, we focus on lab-based studies that allow controlled investigation of causal relationships, rather than population-level epidemiological studies that can detect only associations (Hsiang et al., 2013). Finally, by integrating recent methodological advances in meta-analysis that address issues of small study and publication bias (e.g., PET-PEESE, *p*-curve), and incorporating published and unpublished studies, the present paper can better correct for skewed meta-analytic estimates (Schmidt & Oh, 2016) that are likely to inflate effect size estimates in existing meta-analyses, and thereby provide meta-analytic estimates that are a fairer reflection of the true underlying effect sizes.

To achieve this, we first collated existing experimental and quasi-experimental studies examining temperature effects on prosocial and antisocial behaviour. We then assessed evidence that increases in temperature lead to increases in prosocial and/or antisocial behaviour. In doing so, we considered potential sensory moderators (i.e., whether temperature experience is ambient and on-going, or whether it is haptic and transient) and potential contextual moderators (i.e., whether the valence of the experimental interaction can be construed as positive, negative, or neutral). Finally, we also assessed the impact of publication bias on the included studies. As such, this meta-analysis enables us to synthesise empirical evidence for the key theoretical perspectives (i.e., *heat-facilitates-aggression*, *warmth-primes-prosociality*), and to provide specific recommendations for the future development of the field.

Method

This meta-analysis focuses on four distinct sub-questions: 1) What is the relationship between hotter temperature experience and increases in antisocial behaviour? 2) What is the relationship between hotter temperature experience and increases in prosocial behaviour? 3) Are prosocial and antisocial outcomes equally affected regardless of whether temperature experience is via ambient temperature (i.e., immersive, on-going) or haptic temperature experience (i.e., localised, transient)? and 4) Are temperature effects on behaviour moderated by situational context?

Although the specific protocol was not pre-registered, as far as possible we have followed the guidance from the Cochrane Collaboration's handbook for systematic reviews (Higgins et al., 2019; available at http://handbook.cochrane.org/), which is considered the gold standard for meta-analyses for clinical interventions (Goldacre, 2014). Below, we provide the specific criteria used to select studies for the meta-analyses, while also providing examples of studies that fall beyond the scope of this analysis, for comparison. We also supplement this approach with additional analyses to consider publication bias and other sources of bias. We adopted the Cochrane Handbook's recommendation of using the PICO acronym (Participants, Interventions, Comparisons, Outcomes) to help specify the review question, and select studies for the meta-analysis that meet these clearly-defined following criteria:

- *Participants*: Participants are drawn from the general population (i.e., not clinical samples). Non-human animal subjects are excluded.
- *Interventions*: Interventions consist of studies where the independent variable (IV) is temperature, which is either an experimental manipulation of temperature or an observed and measured IV. In all cases, temperature is related to physical variation in the degree or intensity of heat. As such, manipulations of ambient temperature, physical interactions with objects of different temperatures, and studies conducted at different (but non-manipulated) ambient temperatures meet this criterion. Manipulations that do not meet this criterion, and thus are not included in the meta-analysis, include “imagined” warmth (e.g., Macrae et al., 2013), figurative uses of heat-related terminology (e.g., flavours that are considered “hot” or “cold”, Lewandoski et al., 2012), and paradigms in which temperature (or an estimate of temperature) is an *outcome* measure rather than a manipulation (e.g., IJzerman et al., 2012; Zhong & Leonardelli, 2008). As discussed above, we do not consider epidemiological/archival studies that look at behaviours at a population level over a period of time
- *Comparisons*: The conditions being compared are hot (or warm) vs. cold (or cool) conditions. Some studies additionally included a neutral comparison condition, so the focal comparison may be hot vs. neutral or cold vs. neutral[1].

[1] A brief note on the use of warm and hot temperature conditions: The human thermoneutral zone is in the range of 18-22C, and when temperature goes beyond this point it starts to feel unpleasant for people, in part due to different thermoreceptors responding to different temperature levels. Of course, there are individual differences in what

- *Outcomes*: Outcome measures (dependent variables) will be any quantifiable measures that are an operationalised manifestation of prosocial or antisocial behaviour. Prosocial behaviours are defined as voluntary actions, or thoughts about such actions, that are intended to benefit another person or society as a whole (e.g., Eisenberg et al., 2007). Examples include helping, donating, sharing, and cooperating (Brief & Motowidlow, 1986). Antisocial behaviours are defined as actions, or thoughts about such actions, that harm or lack consideration for the well-being of others or society as a whole (e.g., Berger, 2003). Examples include obstructing others, stealing, being selfish, and sabotaging. Examples of measures that have been used in investigating temperature-behaviour links, but that do not meet the criteria for prosocial or antisocial measures, and are therefore excluded from the meta-analysis, are: extent of concrete language use, perceived distance from an object, ratings of positive/negative affect (e.g., IJzerman & Semin, 2009; Wang, 2017).

To search for papers that met the above criteria, we first used Google Scholar[2] to examine all papers that had cited Williams and Bargh's (2008) paper (537 hits, 02/10/2014), and repeated this search on 01/05/2018, giving an additional 651 hits. Google Scholar was selected, as it provides greater coverage of viable outlets (e.g., PhD dissertations, conference publications) compared to other databases (e.g., Web of Science). Subsequently, we conducted additional searches of Psych File Drawer, the Open Science Framework, and ProQuest PhD database for

---

people experience as a thermoneutral temperature – what is "hot" to one person might be considered neutral to another – with thermoneutral temperatures also dependent on an individual's sex and weight (Kingma et al., 2012). While future studies should take these factors into consideration, the studies included in this paper do not make such distinctions, and it is not possible to readily separate people's experience of "warm" versus "hot". For this reason, our comparisons are based on contrasts between hot OR warm conditions and cold OR neutral conditions. We thank Hans IJzerman for this suggestion.

[2] Use of Google Scholar makes it difficult to make a search process reproducible, but because of the broader coverage offered, especially of grey literature, we took the decision to use Google Scholar in our searches.

applicable studies. The search terms used in Google Scholar (17/10/2014) and Web of Science were as follows: [temperature AND (prosocial OR antisocial OR altruism OR altruistic OR aggression OR aggressive); temperature AND (helping OR hostile OR hostility); temperature AND ("social warmth" OR "social coldness" OR "social cold"; "warm condition" "cool condition" social)]. We also made public calls for papers via the International Social Cognition Network (ISCON) Facebook page, Twitter, and a personal blog post (July 6th, 2015[3]).

Following these searches, the removal of duplicates, and initial screening for inclusion, a long-list of 81 potential articles was created. Where necessary, authors of these articles were contacted to provide additional details of methodology or findings, and to discern if they had conducted any similar studies that may not have been published (e.g., possibly because of null-effects or more recent studies that may not yet be fully written up). Articles were coded using the meta-analytic coding system developed by Lakens et al. (2017). Appendix B shows the flow diagram for the search process, and exclusions at each stage (Moher et al., 2009).

The first author collated a list of the dependent variables (outcome measures) and the other three authors independently rated if they considered the dependent variable to reflect prosocial/altruistic or antisocial/aggressive behaviour, according to the criteria outlined above. Any disagreements were resolved through discussion, with unanimous agreement from authors on the decision to include/exclude each study in the final meta-analysis.

Following Anderson et al. (2000), we also categorised each study according to the valence of the experimental context. Contexts were coded as negative, for example where the participant has a negative social experience, such as being insulted; as positive, where the

[3] Calls for studies were made via Facebook, Twitter, and an Author Blog. We have removed the direct links from the present manuscript to preserve blinding during the review process, but links will be included prior to publication.

participant has a positive social experience, such as receiving a refreshing drink, or as neutral, where there was no additional social or contextual manipulation.

We describe below the full data set and how effect sizes for each study were calculated. Screengrabs of where included effects are extracted from original papers are available here: https://osf.io/xpyzj/?view_only=8b3288219cbb487397a0fdd1c3260d23

Data Preparation and Effect Size Calculation

Eighty-seven effect sizes from 31 articles were identified as potentially eligible for inclusion in the meta-analysis. Six results had insufficient information for computing an effect size (four effects from Boyanowsky, 1999; two effects from van Goozen et al.,'s (1994) first dependent variable). One additional effect (the first helping study in Schneider et al., 1980) had insufficient variance in the dependent variable (in the hot and cold conditions, respectively, 18 of 19 and 21 of 21 participants helped a confederate). Thus, the analysis was based on 80 effect sizes from 30 articles. Some of the effect sizes were dependent (e.g., multiple dependent variables from the same sample); sixty-seven of the 80 effect sizes were independent.

For each study with a continuous dependent variable, we gathered means, standard deviations, and sample sizes and computed standardized mean differences (Hedge's *g*) using the *metafor* package in R (Viechtbauer, 2010). If these raw values were unavailable, we converted *F* values for the main effect of temperature into standardized mean differences using the *compute.es* package in R (Del Re, 2013). If the *F* value for the main effect was unavailable, we used any available *F*, together with the best available estimate of cell means (e.g., taken from figures), to estimate mean squared error for the ANOVA. We then took the square root of this value to obtain an estimate of the pooled standard deviation of the study conditions (Johnson & Eagly, 2000). We then used the estimated means and estimated pooled standard deviation to compute the standardized mean difference. For studies that featured factorial designs with additional factors besides temperature, each simple effect was entered in the meta-analysis individually.

For dichotomous dependent variables, we predominantly used reported proportions and sample sizes to form 2 × 2 tables and compute log odds ratios. These log odds ratios were then converted to a standardized mean difference using the formulas from Borenstein et al., (2009).

One effect was represented only as a correlation coefficient (Kenrick & McFarlane, 1986), and two effects were given as standardized logistic regression coefficients (Lynott et al., 2017). Each of these effects was therefore converted to a standardized mean difference using the formulas in Borenstein et al. (2009).

As noted above, some of the studies contributed multiple effect sizes to the analysis. We handled these dependencies in two ways. First, we computed a multivariate meta-analysis (Olkin & Gleser, 2009), which essentially allows effect sizes to be nested within studies (i.e., the dependence in effect sizes is taken into account). Second, we averaged dependent effect sizes, so that each study contributed only one effect size to the analysis, and then conducted a random-effects meta-analysis. Similar results were obtained from both procedures.

All dependent variables were coded so that higher scores indicated more prosocial responding. In most cases, this meant that studies with antisocial dependent variables needed to be reverse scored prior to analysis, so that these lower scores indicated less prosocial (and more anti-social) responding. Therefore, all effect sizes were calculated such that an effect size greater than zero meant that there was a more prosocial response to hot as compared to cold stimuli. Consequently, proponents of the *warmth-primes-prosociality* hypothesis would predict a significant positive meta-analytic estimate (i.e., greater than zero), whereas proponents of the *heat-facilitates-aggression* hypothesis would predict a significant negative meta-analytic estimate (i.e., less than zero).

We first performed an omnibus analysis with all included studies, and then performed follow-up moderation analyses to assess whether the overall effect varied based on type of dependent variable (prosocial vs. antisocial), social context (positive, neutral, negative), and their interaction. We next considered type of temperature prime (haptic vs. ambient), type of dependent variable (prosocial vs. antisocial), and their interaction. Because of the limited number of effects, we did not model all three moderators in one analysis. Type of dependent variable and type of prime were dummy coded (0 = antisocial/haptic, 1 = prosocial/ambient), and social context was contrast coded with one contrast comparing positive and negative social context vs. a neutral context (neutral = -2, positive = 1, negative = 1) and another contrast comparing a positive to a negative context (neutral = 0, positive = 1, negative = -1). Note that categorizations of study context were straightforward, as experiment descriptions explicitly stated if there was additional positive or negative manipulations, with 100% agreement between authors in assigning these categories.

For the analyses of the type of dependent variable, type of prime, and social context, we also evaluated model fit with Bayesian model comparison, using Bayesian Information Criterion (BIC) to calculate Bayes Factors for the alternative model (BF: Wagenmakers, 2007). A Bayesian model comparison approach allows us to compare the fit of the data under the null hypothesis, compared to the alternative hypothesis (i.e., including temperature-related variables) providing a means to quantify the strength of evidence for and against each model being considered (Dienes & McClatchie, 2018). For example, values of $BF_{10} > 3$ are seen as providing positive support for the alternative hypothesis, values between 0.33 and 3 are in the anecdotal range, while values < 0.33 are seen as providing better support for the null hypothesis (Jeffreys, 1961; Kass & Raftery, 1995).

An issue for any meta-analysis is that publication bias may lead to non-significant results not being in the public domain (Rothstein et al., 2006), and so the meta-analysis may be skewed in favour of significant effects. We have attempted to address this issue by asking authors directly about relevant unpublished data, by including studies published as part of PhD theses or conference presentations, and by including results from registered report replication studies where the outcomes are unbiased, since they are published regardless of the results. Nonetheless, even taking an inclusive approach, we note that only a small proportion of the reported effect sizes come from preregistered studies, and selective reporting in the literature may still have introduced bias to the sample. To address this problem, we conducted additional analyses in the form of PET-PEESE (Precision-Effect Testing–Precision-Effect-Estimate with Standard Error, Stanley and Doucouliagos, 2014), *p*-curve analyses (Simonsohn et al., 2014), and a GRADE risk of bias assessment (Sterne et al., 2019).

PET-PEESE checks for the presence of bias in the sample of included studies by examining the relationship between standard errors and reported effect sizes. When there is no publication bias, there should be no relationship between these two measures, but when the sample is biased (i.e., significant results are more likely to be published), there will be a positive relationship between standard error and effect size (see Van Elk et al., 2015). *p-curve* analysis, on the other hand, identifies the consequences of publication bias by examining the distribution of reported p-values. As such, one does not need to know if publication bias is reflected in the sample, but can infer if this is the case from the p-curve analysis. If the sample is biased, there will be a greater proportion of p-values between .025 and .05 than would be expected by chance (i.e., the distribution of values is left-skewed). However, if the effect in question is larger than zero, and the sample is not biased, then there will be a greater proportion of small p-values ($< .025$. i.e., the distribution will be right-skewed). One of the criteria for *p*-curve analysis is that each *p*-value must be statistically independent from other *p*-values. Thus, where a study provides multiple values, we provide an analysis where only the first *p*-value is included from that study, and a second analysis where only the last *p*-value reported is included.

Finally, to consider the risk of other forms of bias, and evaluate the confidence we can have in the overall body of work (i.e., rated from very low to high), we used the GRADE approach from the Cochrane Handbook (Higgins et al., 2019). GRADE considers whether publication bias is present, as well as other forms of bias (e.g., lack of randomization, no preregistration of analysis plans), inconsistency in effects, indirectness of measures, and imprecision of measures. Combining these evaluations allows us to make a qualitative assessment for a group or subgroup of studies that captures our level of confidence in the estimated effects from Very Low to High levels of confidence.

Table 1 presents a summary of all studies included in the meta-analysis. All data, analysis, and scripts can be accessed on the project's OSF page: https://osf.io/xpyzj/?view_only=8b3288219cbb487397a0fdd1c3260d23. We first report omnibus results for all studies, and subsequently break down the findings for i) prosocial and antisocial outcomes, ii) ambient and haptic prop manipulations, iii) situational context effects. We then report supplementary analysis to examine issues relating to study and publication bias, using PET PEESE, p-curve, and GRADE evaluation.

TABLE 1 ABOUT HERE

Meta-analysis results

The omnibus random-effects multivariate model revealed that the meta-analytic effect size of temperature was not significantly different from zero, g = -0.011, 95% CI: [-0.131, 0.110], SE = .062, Z = -0.171, p = .864, BIC = 133.733. There was significant heterogeneity, $Q(79)$ = 183.638, p < .001, tau = 0.372. The model with averaged dependent effect sizes was very similar, g = -0.007, 95% CI: [-0.127, 0.113], SE = 0.061, Z = -0.108, p = .914, BIC = 111.023. This model also had significant heterogeneity, I^2 = 63.25%, $Q(66)$ = 152.375, p < .001, tau = 0.360. Thus, there was no evidence for an overall effect of temperature on behavioural outcomes across these 80 effect sizes from 67 studies.

We tested whether the model was moderated by type of dependent variable (prosocial vs. antisocial). The moderator was not statistically significant, $Q_{Moderator}$ (1) = 1.016, p = .313, BIC = 136.409 (multivariate model), BF_{10} = 0.262; R^2 = 0.00%, $Q_{Moderator}$ (1) = 0.676, p = .411, BIC = 114.022 (simple model), BF_{10} = 0.223. The average effect size (from the multivariate model) for antisocial studies was g = -0.077, 95% CI: [-0.254, 0.100], whereas for prosocial studies it was g = 0.048, 95% CI: [-0.118, 0.214].

Similarly, we tested whether the model was moderated by type of manipulation (haptic vs. ambient). The moderator was not statistically significant, $Q_{Moderator}$ (1) = 1.09, p = .297, BIC = 137.416, BF_{10} = 0.159 (multivariate model); R^2 = 8.22%, $Q_{Moderator}$ (1) = 3.615, p = .057, BIC = 111.291, BF_{10} = 0.875 (simple model). The average effect size (from the multivariate model) for haptic studies was g = -0.069, 95% CI: [-0.235, 0.098], whereas for ambient studies it was g = 0.021, 95% CI: [-0.117, 0.158].

Type of dependent variable and social context

We examined the moderating effects of type of dependent variable (prosocial vs. antisocial) together with social context (neutral, positive, negative) and their interaction. The test of moderation was non-significant, $Q_{Moderator}$ (5) = 4.393, *p* = .494, BIC = 147.003, BF_{10} = 0.005 (multivariate model); R^2 = 0.00%, $Q_{Moderator}$ (5) = 3.419, *p* = .636, BIC = 124.863, BF_{10} = 0.004 (simple model). Nonetheless, for completeness we estimated effect sizes for each of the six types of studies; results are displayed in Table 2, Figure 1, and Figure 2.

The interaction between the contrast code that compared positive to negative contexts and the prosocial dummy code was not significant, *Z* = 1.008, *p* = .314 (multivariate model) and *Z* = 1.010, *p* = .312 (simple model). Therefore, in contrast to Anderson et al., (2000), we did not observe an effect of temperature on antisocial behaviour in negative social contexts.

Type of dependent variable and type of prime

We next considered whether the method of temperature exposure (ambient vs. haptic) moderated effect sizes, and whether these varying methods had different effects for antisocial vs. prosocial dependent variables. The test of moderation was not significant, $Q_{Moderator}$ (3) = 2.909, *p* = .406, BIC = 143.461, BF_{10} = 0.029 (multivariate model); R^2 = 14.19%, $Q_{Moderator}$ (3) = 7.289, *p* = .063, BIC = 115.359, BF_{10} = 0.512 (simple model). Neither the main effects, nor the interaction between type of dependent variable and type of prime, were statistically significant. Regardless, the effect size estimates (from the multivariate model) were as follows: for ambient, prosocial studies (*k* = 18, *g* = -0.179 [-0.437, 0.078]), haptic, prosocial studies (*k* = 17, *g* = -0.009 [-0.205, 0.187]), ambient, antisocial studies (*k* = 41, *g* = -0.061 [-0.260, 0.138]), and haptic, antisocial studies (*k* = 4, *g* = -0.178 [-0.672, 0.316]). For certain moderator analyses, the small number of studies (*k*) should be noted, and results treated with appropriate caution.

Tests of Bias

To consider possible biases in the meta-analysis studies, we report the meta-analysis funnel plot, PET-PEESE analysis, and summarise the findings of a p-curve, and Risk of Bias analyses.

Figure 3 shows a contour-enhanced funnel plot of the 67 independent effects in the meta-analysis. Visual inspection of the plot reveals a range of effect sizes, including many small/null effects. Studies with prosocial dependent variables tended to have greater power, and in agreement with the analysis above, had somewhat more positive (i.e., warm = good) effects. By contrast, studies with antisocial dependent variables tended to have very imprecise estimates, and effects were somewhat more likely to be negative than positive (i.e., warm = bad). However, there is no obvious asymmetry in the distribution of effect estimates.

PET-PEESE

A PET-PEESE meta-regression was conducted to examine evidence for funnel plot asymmetry (Stanley & Doucouliagos, 2014). We used the formulas developed by Carter and McCullough (2014) for R. PET-PEESE is a pair of weighted-least-squares regressions in which effect sizes are first regressed on their standard errors (the PET phase of the analysis). If the intercept of the regression is statistically significant, indicating evidence for a meta-analytic effect that is different from zero, the analysis is followed up with a second regression in which effect sizes are regressed on sampling variances (the PEESE phase). We did one set of analyses with all 67 independent effect sizes, and additional subgroup analyses for prosocial and antisocial effects. Results are displayed in Table 3. In none of the analyses was the intercept or the funnel plot asymmetry test (i.e., the slope) statistically distinguishable from zero.

Given that a single significant contrast in a complex design may have been enough to publish a study, it is perhaps unsurprising that there does not appear to be much evidence for publication bias overall. As we consider further in the general discussion, it may also be the case that flexibility in the theoretical framework (i.e., hot can be thought to facilitate either prosocial or antisocial behaviour), combined with small samples, makes publishing nearly any study possible.

p-curve

A *p*-curve examines the distribution of *p*-values to understand the evidential value of a set of studies (Simmons et al., 2014a; 2014b). Full details of the p-curve analysis are provided in Appendix A, including the *p*-curve disclosure (also viewable at https://osf.io/d6jqv/?view_only=8b3288219cbb487397a0fdd1c3260d23. Overall, the p-curve analyses suggest that the effects of temperature on antisocial outcomes are weak and lack robustness, while the evidence for effects on prosocial outcomes remains inconclusive.

Risk of Bias and Overall Quality Assessment of included Studies

We used the Cochrane Risk of Bias 2 assessor tool (Sterne et al., 2019) to code for each study's risk of bias related to randomisation procedures, deviations from intended interventions, missing outcome data, measurement of outcomes, and pre-registration of analysis plans. A sample of studies was blind, double-coded, with any discrepancies resolved through discussion. In line with these judgements, the first author coded the remainder of the studies. Figure 4 shows the overall bias by domain.

Insert Figure 4 about here

Overall 65.8% of studies were classified as "high risk", 26.3% as having "some concerns", and the remaining 7.9% as "low risk". The reasons for the small proportion of "low risk" studies is partly due to the fact that the RoB 2 tool is normally applied to studies involving clinical interventions, and include tests of assignment to intervention such as measurements of intentions-to-treat (ITT), which are not currently standard in other empirical domains. Thus, even high-quality registered report studies would not necessarily be classified as "low risk" by the RoB 2's internal algorithm. Nonetheless, many studies show higher risk due to a lack of preregistered analysis plans, having multiple possible outcomes measures, or multiple possible analysis plans (78.9%). In some cases, greater bias is due to possible deviations from the intended outcomes, often due to a lack of blinding, with participant and/or researcher awareness of allocation to conditions (68.4%).

Insert TABLE 4 around Here

We also see differences in overall confidence when considering studies grouped by the ambient/haptic independent variable and by prosocial/antisocial outcomes (Table 4). Studies investigating prosocial outcomes, with haptic temperature manipulations emerge as having Moderate certainty, which is driven by the inclusion of several preregistered studies, and several studies with high statistical power. By contrast, studies with antisocial or prosocial outcomes using ambient temperature manipulations, and studies with antisocial outcomes and haptic manipulations give rise to Very Low certainty of results. This reduced certainty stems from differences in populations used, generally small sample sizes, and considerable variability in effect sizes. Studies examining Antisocial outcomes using Haptic temperature manipulations are also summarised as having Very Low certainty due to including only a small number of small-sample studies, that also show high variability in observed effect sizes.

General Discussion

We assessed the evidence that temperature variation is linked to changes in prosocial and antisocial behaviour. In a primary meta-analysis, we find no overall effect of temperature on behaviour, no effect of temperature on prosocial or antisocial outcomes when considered separately, no effect of ambient or haptic temperature manipulations, and no reliable effects of (or interactions with) social context. Follow-up PET-PEESE analysis did not provide evidence for publication bias, but there may be other reasons why published studies may overestimate likely effect sizes, which we return to below. Using *p*-curve analyses, we found only weak evidential support for temperature effects on antisocial outcomes, and no evidential support for temperature effects on prosocial outcomes, although the latter tended to have larger sample sizes and therefore narrower confidence intervals around effect size estimates (see e.g., Figure 2). Following a Risk of Bias assessment, we generally observed Very Low certainty around the outcomes of most studies, although studies with Prosocial outcomes and haptic temperature manipulations emerged with moderate certainty outcomes (e.g., Table 4). Finally, it was striking to see that estimated Bayes Factors did not indicate support for temperature-based models; instead, all Bayes Factors were either in the inconclusive range, or indicated better support for null models (i.e., those without temperature-related variables). While any individual analytical approach will have advantages and disadvantages, the convergent evidence provided by the combination of the primary meta-analysis, PET-PEESE, p-curve, and Bayes Factors, all pointing in the same direction, means that we can have reasonable confidence in the conclusion that that effects of temperature on behaviour are not well-supported by existing data.

This is the first meta-analysis to contrast temperature effects on both positive (prosocial) and negative (antisocial) behavioural outcomes. This work is important since temperature represents a fundamental aspect of people's environmental experience, and one that has been increasingly under the spotlight due to climate change and global warming. Furthermore, the findings present a challenge for existing frameworks of temperature-behaviour links since we found that both the *heat-facilitates-aggression* and *warmth-primes-prosociality* perspectives are generally lacking in evidential support. Taking these findings into account, we reconsider both theoretical perspectives, before offering some suggestions for the field, both in terms of enhancing methodological rigour and in making theoretical progress.

Where to for the *warmth-primes prosociality* view?

We observed no reliable effects between higher temperature experience and increased prosocial responding, with a meta-analytic effect size of g = 0.048 [95% CI = -0.118, 0.214]. With such weak effects, it seems premature to try and determine what cognitive and physiological mechanisms might drive temperature-behaviour links. If there is no genuine causal link between temperature experience and behavioural outcomes, we may simply find ourselves attempting to impose structure on noise. Nonetheless, we may be in a better position to consider and reject proposed mechanisms in the literature, based on the lack of robust effects.

For example, proponents of the *warmth-primes-prosociality* approach have suggested an associative priming mechanism to account for previously observed links between interactions with warm (compared to cool) objects (e.g., Williams & Bargh, 2008) and increased prosocial responding. Furthermore, Williams (2008) explicitly suggests that only transient haptic experience should impact prosocial responding, and ambient temperature should not (see also Lynott, et al., 2017). However, the lack of effect of haptic temperature experience (meta-analytic effect size for haptic prosocial studies: g = -0.009 [95% CI = -0.205, 0.187]), and lack of any effect on prosocial responding weakens the suggestion that associative priming is a viable candidate to explain temperature-behaviour links.

Perhaps the lack of a robust effect should be unsurprising, as such a mechanism would require that higher temperature experiences are diagnostically associated with prosocial responses over non-prosocial responses. That is, there should be greater cue validity between higher temperature experiences and prosocial outcomes than the cue validity between higher temperature experience and other outcomes. Considering a thermoregulation perspective, it is certainly true that early experience between caregivers and infants is critical for infant development (Bowlby, 1969; Harlow, 1958), but infants also experience warmth/heat that is not necessarily associated with positively valenced outcomes. For example, infants may experience warmth, but also distress when being held by someone other than a primary caregiver. Infants experience higher bodily temperatures during fever and illness, but again with more negative associations. Beyond early development, interactions with warmth are likely to become even more heterogeneous, so that increasing physical warmth becomes an even less diagnostic cue for prosocial responding, as any associative links become further diluted with more diverse experiences over different developmental pathways. While associative mechanisms may have a role to play in temperature-behaviour links, more stringent experimental studies are needed to support such a framework.

Where to for the *heat-facilitates-aggression* view?

Only in the *p-curve* analysis did we observe weak support for the *heat-facilitates-aggression* view, but this analysis includes only seven statistically significant results. The results of the multivariate and the simple meta-analyses do not support an effect of temperature on antisocial outcomes, and the Grade Risk of Bias analysis reveals Outcome Certainty as being

Very low for these studies. While Anderson and colleagues (2000) found differences between social context (positive, negative, neutral) on aggressive outcomes, we observed no such support in our analyses. Indeed, the Bayes Factors for these models suggest that the null hypothesis is better supported by the data.

An important question for the *heat-facilitates-aggression* perspective is to ask why population-based studies appear to support temperature-behaviour links to a far greater extent than lab-based studies. Meta-analyses of population-based research strongly suggests an association between temperature and negative societal outcomes, all the way from horn honking to civil wars (Hsiang et al., 2013). However, because of the difficulties in determining causal links in many of these studies, it may be that temperature effects are a proxy effect for some other factor, be it another climatic measure, or some unmeasured socio-economic factor (e.g., unemployment, lack of resources, population density). For example, warmer temperatures often mean increased likelihood of interpersonal interactions, which could increase the probability of antisocial outcomes (e.g., Rotton & Cohn, 2004). It may also be that meta-analyses of population-based studies overestimate effect sizes due to factors such as publication bias. For example, Hsiang et al's 2013 meta-analysis contains no unpublished studies (which are more likely to contain non-significant effects). Furthermore, the level of flexibility available to researchers in specifying statistical models (e.g., Gelman & Loken, 2013), can also increase the likelihood of obtaining false-positive results (Simmons et al., 2011), thereby inflating meta-analytic estimates of temperature effects on population-level antisocial outcomes.

Alternatively, it may be that there are problems with lab-based studies that do not capture important aspects of field and population-based studies, and therefore underestimate the true strength of temperature-behaviour links.

Consider that beyond human population-level studies, there is evidence of a link between temperature change and behaviour in animal studies too. For example, warmer water conditions have been linked to increases in aggressive behaviour amongst species of tropical fish, such as those found in Lake Tanganyika in East Africa (Kua et al., 2020. See also, Biro et al., 2010). However, Van Lange and colleagues (2017) caution against making strong claims between temperature differences (e.g., due to climate change) and animal behaviour, due to a lack of good empirical evidence. As a counterpoint, work by IJzerman and colleagues (IJzerman & Hogerzeil, 2017) describe how thermoregulation in humans and animals, can be intrinsically linked to *prosocial* behaviours. For example, social grooming in vervet monkeys was associated with better protection against rapid external temperature changes (McFarland et al., 2016). While animal studies may provide insights for research on human behaviour, the picture is far from clearcut, with plenty of work yet to be done in this domain.

Nonetheless, we should ask whether the ecological validity of laboratory-based tasks is too low. For example, is it realistic to have participants choose prison sentences for accused targets? Or to make a forced choice between selecting a gift for friends or oneself? These are subjective judgements, but perhaps researchers should be more cognisant of having realistic and ecologically valid tasks to narrow the explanatory gap between individual-focused, lab-based studies and population-focused, epidemiological studies.

Where do we go from here?

At the outset, we highlighted the problems of psychology in terms of reproducibility, and how studies of temperature-behaviour links were not immune to the issues surrounding reproducibility and the “replication crisis.” In this final section, we take stock of our findings and offer some guidance for researchers in this and cognate areas, who wish to build towards a more cumulative psychological science, where issues of low power studies and publication bias are consigned to the trashcan of history. Some points relate specifically to investigations of the effect of temperature on human behaviour, while others speak to broader issues.

First, those investigating temperature-behaviour links should adopt best-practices from open science, ensuring as far as possible that studies are high-powered, design and analysis plans are preregistered, and that data is shared as openly as possible (see Chambers, 2019, and IJzerman et al., 2021 for similar recommendations). We see movement towards larger sample sizes in this area (e.g., IJzerman, Lindenberg et al., 2018; IJzerman, Neyroud et al., 2018; Lynott et al., 2014; Vergara et al., 2019; Wittmann et al., 2021; Sarda et al., 2020), use of preregistration, and data sharing, but there is scope for greater uptake of these elements.

Study preregistration also provides an opportunity to incorporate moderating factors of theoretical interest *a priori*, rather than introducing moderators after-the-fact or in a data-driven manner, which can lead to elevation of false-positive findings (Simmons et al, 2011). Many moderators have been suggested for temperature-behaviour links, including attachment (IJzerman et al., 2013), prior anger (Baron & Bell, 1975), and fear of negative evaluation (Fay & Maner, 2014), but few have been subject to study preregistration. Researchers could also incorporate other plausible moderators that have not yet been properly tested such as country of origin, in-group/out-group status, individual differences in hormonal profiles and responses, male-to-female ratio in the study sample, whether experimental or natural groups are used, and even individual differences in response to temperature (i.e., personal “real feel” measures). For example, a Hot condition for a study run in central Texas is likely very different to a Hot condition for a study conducted in Northern Norway. What’s more, people living in different temperature environments may respond in different ways due to acclimatisation, and what feels extremely hot and aversive for one individual (or group) may feel moderate and comfortable for another. Thus, considering the thermal comfort of participants (e.g., Oleson, 1982), beyond the actual measured temperature could be a way of accounting for important group and individual differences in our understanding of temperature effects**.** Supporting this view, IJzerman and colleagues (2021) recommend that researchers record a range of variables, which could then be deposited in a *shared* database, which would allow for a rich examination of potential explanatory factors.

Answering the questions of how and when temperature impacts behaviour will be greatly helped by expanding temperature ranges used in studies, by considering antisocial and prosocial behaviour broadly construed, and by using within-participant designs to further enhance statistical power. Currently, most lab-based studies use factorial designs, where temperatures are binned into cold, neutral, and hot conditions for example. Extending the range of temperatures considered will help answer the question of whether temperature is linearly (e.g., Larrick et al., 2011) or non-linearly (Hsiang et al., 2013; Rotton & Cohn, 2004) related to behavioural outcomes, which is not possible with traditional factorial designs. Such studies would also help overcome issues around the specification of what is meant by “hot” or “cold” temperature conditions, and at what temperature ranges theoretical predictions can be reliably applied. Of course, including wider temperature ranges increases study costs, and so a balance must be struck. If available resources means that researchers run more modest studies that is fine; provided there is full transparency, such studies can meaningfully contribute to the literature.

To get a fuller picture of the behavioural ramifications of varying temperature experience, researchers should also take a broad perspective, and if possible, directly contrast prosocial and antisocial outcomes for the same independent measures, while also considering the ecological validity of tasks and measures. Some studies do well in having pseudo-realistic measures, such as Kenrick and McFarlane's (1986) study of horn-honking, whereas others have used relatively unrealistic scenarios, such as applying noise blasts to other participants.

Finally, we feel it is critical for researchers to provide clear conceptual and operational definitions for independent and dependent measures. We noted in the introduction that differences in defining what constituted "conflict" lead to contrasting conclusions about whether temperature changes influence societal stability and violence (Buhaug, 2010). In social psychology, there may be issues with how labels are used and how dependent measures are operationalised. For example, Asch (1946) and Fiske and colleagues (2007), identify "warmth" as a key dimension in making judgements and forming impressions of others, suggesting a "primacy-of-warmth" over other dimensions, such as "competence". However, the concept of "warmth" seemingly encapsulates a broad range of other concepts including trust, morality, honesty, generosity, as well as other traits (Fiske et al., 2007). Labels with such broad scope make it difficult to make concrete claims about links between temperature and behaviour, as the behavioural outcomes become so diffuse, which is problematic for testing a *warmth-primes-prosociality* perspective. Morality and interpersonal warmth can hardly be considered the same constructs, and they should not be conflated, to avoid the jingle fallacy (see e.g., Kelley, 1927; Marsh, 1994). Should higher temperature experiences equally lead to increases in honesty, generosity, morality, and so on? Although an open empirical question, it seems unlikely, and an a priori theoretically weak position to start from. Researchers should aim to be specific about delineating the behaviours that are likely to be influenced by temperature, and they should clearly define how such behaviours are measured, to allow meaningful interpretation of observed effects. Providing greater clarity around measures and manipulations will also facilitate future meta-analyses, enabling researchers to easily assess studies against strict inclusion criteria.

From even this short list of suggestions, it is obvious that there is scope for much deeper examination of the question of how temperature experience might influence behaviour, but also that researchers can do much more to strengthen their case by designing and running better, highly powered studies, and by adopting open-science practices that aid in the collective goal of achieving a cumulative science of temperature-behaviour links.

Conclusions

Following a review of the literature and a meta-analysis considering the effects of temperature experience on behaviour, we find that there is no clear support for two commonly adopted theoretical perspectives: *warmth-primes-prosociality* or *heat-facilitates-aggression*. However, we identify several routes for building on this research and working towards a better understanding of temperature-behaviour links. Given what we know of the potential for temperature effects at a geographical and societal scale, it is incumbent on psychologists to seek a better understanding at the level of the individual.

References

Anderson, C. A. (1987). Temperature and aggression: Effects on quarterly, yearly, and city rates of violent and nonviolent crime. *Journal of Personality and Social Psychology, 52,* 1161-1173.

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

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

Anderson, C. A. & Anderson, D.C. (1984). Ambient temperature and violent crime: Tests of the linear and curvilinear hypotheses. *Journal of Personality and Social Psychology, 46*, 91-97.

Anderson, C. A. & Anderson, D.C. (1996). Violent crime rate studies in philosophical context: A destructive testing approach to heat and southern culture of violence effects. *Journal of Personality and Social Psychology, 70,* 740-756.

Anderson, C. A., Anderson, K. B., & Deuser, W. E. (1996). Examining an affective aggression framework weapon and temperature effects on aggressive thoughts, affect, and attitudes. *Personality and Social Psychology Bulletin*, *22*(4), 366-376.

Anderson, C. A., Anderson, K. B., Dorr, N., DeNeve, K. M., & Flanagan, M. (2000). Temperature and aggression. *Advances in experimental social psychology, 32,* 63-133.

Anderson, C. A., & Bushman, B. J. (2002). The effects of media violence on society. *Science*, *295*(5564), 2377-2379.

Anderson, C. A., Bushman, B. J., & Groom, R. W. (1997). Hot years and serious and deadly assault: empirical tests of the heat hypothesis. *Journal of personality and social psychology*, *73*(6), 1213.

Anderson, C. A., & DeLisi, M. (2011). Implications of global climate change for violence in developed and developing countries. Chapter in J. Forgas, A. Kruglanski, & K. Williams (Eds.), *The Psychology of Social Conflict and Aggression*. (pp. 249-265). New York: Psychology Press.

Anderson, C. A., Deuser, W. E., De Neve, K. (1995). Hot temperatures, hostile affect, hostile cognition, and arousal: tests of a general model of affective aggression. *Personality and Social Psychology Bulletin, 21*, 434–48.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal Psychology, 41,* 258–290.

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

Baron, R. A. (1972). Aggression as a function of ambient temperature and prior anger arousal. *Journal of personality and social psychology*, *21*(2), 183.

Baron, R. A. (1972). Aggression as a function of ambient temperature and prior anger arousal. *Journal of Personality and Social Psychology, 21(2),* 183-189.

Baron, R. A. (1976). The reduction of human aggression: A field study of the influence of incompatible reactions. *Journal of Applied Social Psychology, 6,* 260–274.

Baron, R. A., & Bell, P. A. (1975). Aggression and heat: Mediating effects of prior provocation and exposure to an aggressive model. *Journal of Personality and Social Psychology, 31(5),* 825-832.

Baron, R. A., & Bell, P. A. (1976). Aggression and heat: The influence of ambient temperature, negative affect, and a cooling drink on physical aggression. *Journal of Personality and Social Psychology, 33(3),* 245-255.

Beckes, L., Ijzerman, H. & Tops, M. (2015). Toward a radically embodied neuroscience of attachment and relationships. Frontiers in Human Neuroscience, http://dx.doi.org/10.3389/fnhum.2015.00266

Belkin, L. Y., & Kouchaki, M. (2017). Too hot to help! Exploring the impact of ambient temperature on helping. *European Journal of Social Psychology*, *47*(5), 525-538.

Bell, P. A., & Baron, R. A. (1974). Environmental influences on attraction: Effects of heat, attitude similarity, and personal evaluations. *Bulletin of the Psychonomic Society, 4(5),* 479-481.

Bell, P. A., & Baron, R. A. (1976). Aggression and Heat: The Mediating Role of Negative Affect1. *Journal of Applied Social Psychology, 6(1),* 18-30.

Bell, P. A., & Baron, R. A. (1977). Aggression and ambient temperature: The facilitating and inhibiting effects of hot and cold environments. *Bulletin of the Psychonomic Society, 9(6),* 443-445.

Berger, K. S. (2003). *The developing person through childhood and adolescence*. Macmillan.

Biro, P. A., Beckmann, C. & Stamps, J. A. Small within-day increases in temperature affects boldness and alters personality in coral reef fish. *Proceedings of the Royal Society B: Biological Sciences,* **277**, 71–77.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley.

Bowlby, J. (1969). *Attachment and loss*. London: Hogarth Press.

Boyanowsky, E. (1999). Violence and aggression in the heat of passion and in cold blood: The Ecs-TC syndrome. *International Journal of Law and psychiatry*, *22*(3-4), 257-271.

Brief, A. P., & Motowidlow, S. J. (1986). Prosocial organizational behaviours. *Academy of management Review*, *11*(4), 710-725.

Buhaug, H. (2010). Climate not to blame for African civil wars. *Proceedings of the National Academy of Sciences*, *107*(38), 16477-16482.

Buhaug, H., Nordkvelle, J., Bernauer, T., Böhmelt, T., Brzoska, M., Busby, J. W., ... & Goldstone, J. A. (2014). One effect to rule them all? A comment on climate and conflict. *Climatic Change*, *127*(3-4), 391-397.

Burke, M.B., Miguel, E., Satyanath, S., Dykema, J.D. & Lobell, D.B. (2009). Warming increases the risk of civil war in Africa. *Proceedings National Academy of Sciences, 106*, 20670-20674.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365.

Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology,* 5:823. doi: 10.3389/fpsyg.2014.00823

Chabris, C.F., Heck, P.R., Mandart, J., Benjamin, D.J., & Simons, D.J. (2019). No evidence that experiencing physical warmth promotes interpersonal warmth: Two failures to replicate Williams and Bargh (2008*). Social Psychology, 50,*127–132.

Chambers, C. (2019). What's next for Registered Reports? *Nature*, 573, 187-189

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review, 44,* 588-608.

Del Re, A. C. (2013). *compute.es*: Compute effect sizes. R package version 0.2-2. http://cran.r-project.org/web/packages/compute.es

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5: 781*. doi: 10.3389/fpsyg.2014.00781

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, *25*(1), 207-218.

Donnellan, M. B., Lucas, R. E., & Cesario, J. (2015). On the association between loneliness and bathing habits: Nine replications of Bargh and Shalev (2012) Study 1. *Emotion*, *15*(1), 109.

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2007). Prosocial development. *Handbook of Child Psychology*, *3, 646-718*.

Fay, A. J., & Maner, J. K. (2014). When does heat promote hostility? Person by situation interactions shape the psychological effects of haptic sensations. *Journal of Experimental Social Psychology, 50,* 210-216.

Fay, A. J., & Maner, J. K. (2015). Embodied effects are moderated by situational cues: Warmth, threat, and the desire for affiliation. *British Journal of Social Psychology*, 54(2), 291-305.

Field, S. (1992). The effect of temperature on crime. *The British journal of criminology, 32*(3), 340-351.

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.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*. Retrieved from https://osf.io/n3axs

Gockel, C., Kolb, P. M., & Werth, L. (2014). Murder or not? Cold temperature makes criminals appear to be cold-blooded and warm temperature to be hot-headed. *PloS one*, *9*(4), e96231.

Goldacre, B. (2014). *Bad pharma: how drug companies mislead doctors and harm patients*. Macmillan.

GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. McMaster University, 2020 (developed by Evidence Prime, Inc.). Available from gradepro.org

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

Hewstone, M. E., Stroebe, W. E., & Stephenson, G. M. E. (1996). *Introduction to social psychology: A European perspective*. Blackwell Publishing.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Hsiang, S. M., Burke, M., & Miguel, E. (2013). Quantifying the influence of climate on human conflict. *Science*, *341*(6151), 1235367.

Hsiang, S. M., Meng, K. C. & Cane, M. A. (2011). Civil conflicts are associated with the global climate. *Nature*, *476*, 438-441.

Hu, C. P., Yin, J. X., Lindenberg, S., Dalğar, İ., Weissgerber, S. C., Vergara, R. C., ... & Hadi, R. (2019). Data from the Human Penguin Project, a cross-national dataset testing social thermoregulation principles. *Scientific data*, *6*(1), 32.

Huang, X. I., Zhang, M., Hui, M. K., & Wyer Jr, R. S. (2014). Warmth and conformity: The effects of ambient temperature on product preferences and financial decisions. *Journal of Consumer Psychology*, *24*(2), 241-250.

IJzerman, H., Gallucci, M., Pouw, W. T., Weiβgerber, S. C., Van Doesum, N. J., & Williams, K. D. (2012). Cold-blooded loneliness: Social exclusion leads to lower skin temperatures. *Acta psychologica*, *140*(3), 283-288.

IJzerman, H., Hadi, R., Coles, N. A., Paris, B., Sarda, E., Fritz, W., … Ropovik, I. (2021, February 26). Social Thermoregulation: A Meta-Analysis. https://doi.org/10.31234/osf.io/fc6yq

IJzerman, H., & Hogerzeil, L. J. (2017). People as penguins: Thermoregulation as part of the human essence. *Oxford Handbook of Human Essence. Oxford University Press. DOI: https://doi. org/10.31234/osf. io/zrtg6*.

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

IJzerman, H., Coan, J. A., Wagemans, F. M., Missler, M. A., van Beest, I., Lindenberg, S., & Tops, M. (2015). A theory of social thermoregulation in human primates. *Frontiers in Psychology, 6,* 454.

Ijzerman, H., Gallucci, M., Pouw, W. T. J. L., Weissgerber, S. C., van Dosum, N. J., & Williams, K. D. (2012). Cold-blooded loneliness: Social exclusion leads to lower skin temperatures. Acta Psychologica, 140, 283-288.

IJzerman, H., Karremans, J. C., Thomsen, L., & Schubert, T. W. (2013). How attachment styles modulate communal cues of physical warmth. *Social Psychology*, *44*(2), 160-166.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Jeffreys, H. (1961). *The theory of probability.* OUP Oxford.

Johnson, B. T., & Eagly, A. (2000). Quantitative synthesis of social psychological research. In H. T. Reis & C. M. Judd (Eds.) *Handbook of research methods in social and personality psychology* (pp. 496-528). Cambridge, UK: Cambridge University Press.

Kang, Y., Williams, L. E., Clark, M. S., Gray, J. R., & Bargh, J. A. (2011). Physical temperature effects on trust behaviour: the role of insula. *Social Cognitive and Affective Neuroscience, 6,* 507-515.

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

Kelley, C. P., Mohtadi, S., Cane, M. A., Seager, R., & Kushnir, Y. (2015). Climate change in the Fertile Crescent and implications of the recent Syrian drought. *Proceedings of the National Academy of Sciences, 112(11)*, 3241-3246.

Kelley, T.L. (1927), *The Interpretation of Educational Measurements*, New York: World Book.

Kenrick, D. T., & MacFarlane, S. W. (1984). Ambient temperature and horn-honking: A field study of the heat/aggression relationship. *Environment and Behaviour, 18,* 179-191.

Kingma, B., Frijns, A., & van Marken Lichtenbelt, W. (2012). The thermoneutral zone: implications for metabolic studies. *Front Biosci (Elite Ed)*, *4*, 1975-1985.

Koch, A., Imhoff, R., Unkelbach, C., Nicolas, G., Fiske, S., Terache, J., ... & Yzerbyt, V. (2020). Groups' warmth is a personal matter: Understanding consensus on stereotype dimensions reconciles adversarial models of social evaluation. *Journal of Experimental Social Psychology*, *89*, 103995.

Kolb, P., Gockel, C., & Werth, L. (2012). The effects of temperature on service employees' customer orientation: An experimental approach. *Ergonomics*, *55*(6), 621-635.

Krahé, B. (2013). *The social psychology of aggression*. Psychology Press.

Kua, Z. X., Hamilton, I. M., McLaughlin, A. L., Brodnik, R. M., Keitzer, S. C., Gilliland, J., ... & Ludsin, S. A. (2020). Water warming increases aggression in a tropical fish. *Scientific reports*, *10*(1), 1-13.

Lakens, D., & Evers, E. R. (2014). Sailing from the seas of chaos into the corridor of stability practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science, 9(3)*, 278-292.

Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC psychology*, *4*(1), 24.

Lakens, D., Page-Gould, E., van Assen, M. A. L. M., Spellman, B., Schönbrodt, F. D., Hasselman, F., … Scheel, A. M. (2017, March 31). Examining the Reproducibility of Meta-Analyses in Psychology: A Preliminary Report. https://doi.org/10.31222/osf.io/xfbjf

Larrick, R.P., Timmerman, T.A., Carton, A.M., & Abreyva, J. (2011). Temper, temperature, and temptation: Heat-related retaliation in baseball. *Psychological Science, 22*, 423-428.

Leander, N. P., Chartrand, T. L., & Bargh, J. A. (2012). You give me the chills: Embodied reactions to inappropriate amounts of behavioural mimicry. *Psychological Science, 23*, 772-779.

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

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 & Bargh (2008, Science). *Social Psychology, 45*, 216-222.

McCarthy, R. J. (2014). Close replication attempts of the heat priming-hostile perception effect. *Journal of Experimental Social Psychology*, *54*, 165-169.

McFarland, R., Henzi, S. P., Barrett, L., Wanigaratne, A., Coetzee, E., Fuller, A., ... & Maloney, S. K. (2016). Thermal consequences of increased pelt loft infer an additional utilitarian function for grooming. *American Journal of Primatology*, *78*(4), 456-461.

Macrae, C. N., Raj, R. S., Best, S. B., Christian, B. M., & Miles, L. K. (2013). Imagined sensory experiences can shape person perception: It's a matter of visual perspective. *Journal of Experimental Social Psychology*, *49*(3), 595-598.

Marsh, H. W. (1994). Sport motivation orientations: Beware of jingle-jangle fallacies. *Journal of Sport and Exercise Psychology*, *16*(4), 365-380.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med*, *6*(7), e1000097.

NASA. NASA, NOAA Data Show 2016 Warmest Year on Record Globally 2017 [updated Jan. 18, 2017. Available from: https://www.giss.nasa.gov/research/news/20170118/

Nauts, S., Langner, O., Huijsmans, I., Vonk, R., & Wigboldus, D. H. (2014). Forming impressions of personality. *Social Psychology, 45,* 153-163.

Nosek, B. A., & Lakens, D. (2014). A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137-141.

Olesen, B. W. (1982). *Thermal comfort* (Vol. 2). Bruel & Kjær.

Olkin, I., & Gleser, L. (2009). Stochastically dependent effect sizes. *The handbook of research synthesis and meta-analysis*, 357-376.

Pakes, F. J. (2009). *Criminal psychology.* Routledge.

Palamarek, D. L. & Rule, B. G. (1979). The effects of ambient temperature and insult on the motivation to retaliate or escape. *Motivation and Emotion, 3,* 83-92.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in medicine, 26(25)*, 4544-4562.

Peterson, T. C., Baringer, M. O., Thorne, P. W., Menne, M. J., Kennedy, J. J., Christy, J., ... & Zhang, L. (2009). State of the climate in 2008. *Bulletin of the American Meteorological Society, 90*, S1–S196.

Plotnik, R., & Kouyoumdjian, H. (2013). *Introduction to psychology*. Cengage Learning.

Ranson, M. (2014). Crime, weather, and climate change. *Journal of Environmental Economics and Management*, *67*, 274-302.

Reifman, A. S., Larrick, R. P., & Fein, S. (1991). Temper and temperature on the diamond: The heat-aggression relationship in major league baseball. *Personality and Social Psychology Bulletin*, *17*(5), 580-585.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons.

Rotton, J., & Cohn, E. G. (2000). Violence is a curvilinear function of temperature in Dallas: A replication. *Journal of Personality and Social Psychology*, *78,* 1074-1081.

Rotton, J., & Cohn, E.G. (2004). Outdoor temperature, climate control, and criminal assault: The spatial and temporal ecology of violence. *Environment & Behaviour, 36,* 276–306.

Schilder, J. D., IJzerman, H., & Denissen, J. J. (2014). Physical warmth and perceptual focus: A replication of IJzerman and Semin (2009). *PloS one*, *9*(11), e112772.

Schmidt, F. L., & Oh, I. S. (2016). The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else? *Archives of Scientific Psychology*, *4*(1), 32.

Schneider, F. W., Lesko, W. A., & Garrett, W. A. (1980). Helping behaviour in hot, comfortable, and cold temperatures: A field study. *Environment and Behaviour*, *12*(2), 231-240.

Shariff, A. F., Willard, A. K., Andersen, T., & Norenzayan, A. (2016). Religious priming: A meta-analysis with a focus on prosociality. *Personality and Social Psychology Review*, *20*(1), 27-48.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). p-Curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143,* 534-547.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). p-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9,* 666-681.

Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological bulletin*, 144, 1325-1346.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods, 5*, 60-78.

Sterne, J. A., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., ... & Emberson, J. R. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ, 366*.

Storey, S., & Workman, L. (2012). The effects of temperature priming on cooperation in the iterated prisoner's dilemma. Evolutionary Psychology: An International Journal of Evolutionary *Approaches to Psychology and Behaviour, 11,* 52-67.

Szymkow, A., Chandler, J. J., Ijzerman, H., Parzuchowski, M., & Wojciszke, B. (2013). Warmer hearts, warmer rooms: How positive communal traits increase estimates of ambient temperature. *Social Psychology, 44*(2), 167-176.

Tol, R. S., & Wagner, S. (2010). Climate change and violent conflict in Europe over the last millennium. *Climatic Change*, *99*(1-2), 65-79.

Van Acker, B., Kerselaers, K., Pantophlet, J., & IJzerman, H. (2015). Homely thermoregulation: How physical coldness makes an advertised house a home. Available at SSRN 2576421.

Van Elk, M., Matzke, D., Gronau, Q., Guang, M., Vandekerckhove, J., & Wagenmakers, E. J. (2015). Meta-analyses are no substitute for registered replications: A skeptical perspective on religious priming. *Frontiers in Psychology*, *6*, 1365.

Van Goozen, S. H., Frijda, N. H., Kindt, M., & Van Poll, N. E. D. (1994). Anger proneness in women: Development and validation of the Anger Situation Questionnaire. *Aggressive Behaviour*, *20*(2), 79-100.

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). *The Behavioral and brain sciences*, *40*, e75.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the *metafor* package. *Journal of Statistical Software, 36,* 1-48.

Vrij, A., Van der Steen, J., & Koppelaar, L. (1994). Aggression of police officers as a function of temperature: An experiment with the Fire Arms Training System. *Journal of Community & Applied Social Psychology, 4,* 365-370.

Wang, X. (2017). An Empirical Study of the Impacts of Ambient Temperature on Risk Taking. *Psychology*, *8*(07), 1053.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review, 14,* 779–804.

Willams, L. E. (2008). *Metaphor priming; How basic perceptual experiences nonconsciously shape evaluations, decisions, and emotional experiences*. Yale University.

Williams, L. E. & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322,* 306–307.

Wortman, J., Donnellan, M. B., & Lucas, R. E. (2014). Can physical warmth (or coldness) predict trait loneliness? A replication of Bargh and Shalev (2012). *Archives of Scientific Psychology, 2*(1), 13.

Zhang, D.D., Brecke, P., Lee, H.F., He, Y.Q., Zhang, J. (2007). Global climate change, war, and population decline in recent human history. *Proceedings National Academy of Sciences, 104,* 19214–19219.

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

Appendix A – Details of *p-curve* analysis

The full p-curve disclosure and p-curve output tables are available at: https://osf.io/d6jqv/?view_only=8b3288219cbb487397a0fdd1c3260d23

There are several features of *p*-curve that make it different from the other meta-analytic techniques utilised above. First, *p*-curve is restricted to statistically significant predicted results, meaning only effects that were predicted by the authors are included in the analysis. Any significant effects in the "wrong" (non-predicted) direction are excluded. Because the authors of the antisocial studies predict effects in the opposite direction of the prosocial studies, this means that prosocial and antisocial *p*-curves should be conducted separately. Also, *p*-curve can only accept independent *p*-values, and unlike traditional meta-analysis, dependent effects cannot be aggregated. Instead, analysts are instructed to select the first reported test (i.e., the primary hypothesis test) for the *p*-curve analysis. Secondary hypothesis tests can be substituted in later robustness tests. The *p*-curve disclosure table is linked to in Appendix A, and can be viewed at https://osf.io/d6jqv/?view_only=8b3288219cbb487397a0fdd1c3260d23.

Thus, we conducted two *p*-curves (one prosocial and one antisocial), using only significant effects in the predicted direction. Of 36 independent antisocial effects, seven effects (eight effects in the robustness test) were statistically significant. The antisocial analyses indicated that the studies contain evidential value (Main test: $Z = -2.49$, $p = .0063$; Robustness test: $Z = -2.15$, $p = .0159$). However, the cumulative meta-analysis indicated that after the smallest *p*-value was dropped, the test of evidential value was no longer statistically significant. *p*-curve estimated post-hoc power to be 61% (49% in the robustness analysis). Of 24 prosocial effects, three effects were statistically significant. The *p*-curve analysis was inconclusive, with no evidence for evidential value, low power, or *p*-hacking (all tests $p > .05$). Given the small number of significant effects, it was not possible to incorporate social context into the *p*-curve analysis.

Appendix B

Flow diagram for study/effect selection process. Note that although there were a large number of possible items returned (>100,000), Google Scholar automatically removes many duplicates. and the actual returned number of results is much lower, stopping after the 50 pages of Google Scholar Hits (~5000 papers). Similarly, for Web of Science, relevance of returned articles diminished rapidly (normally within ~200 papers), and searching was halted when a 5000 paper threshold was reached for any particular search.

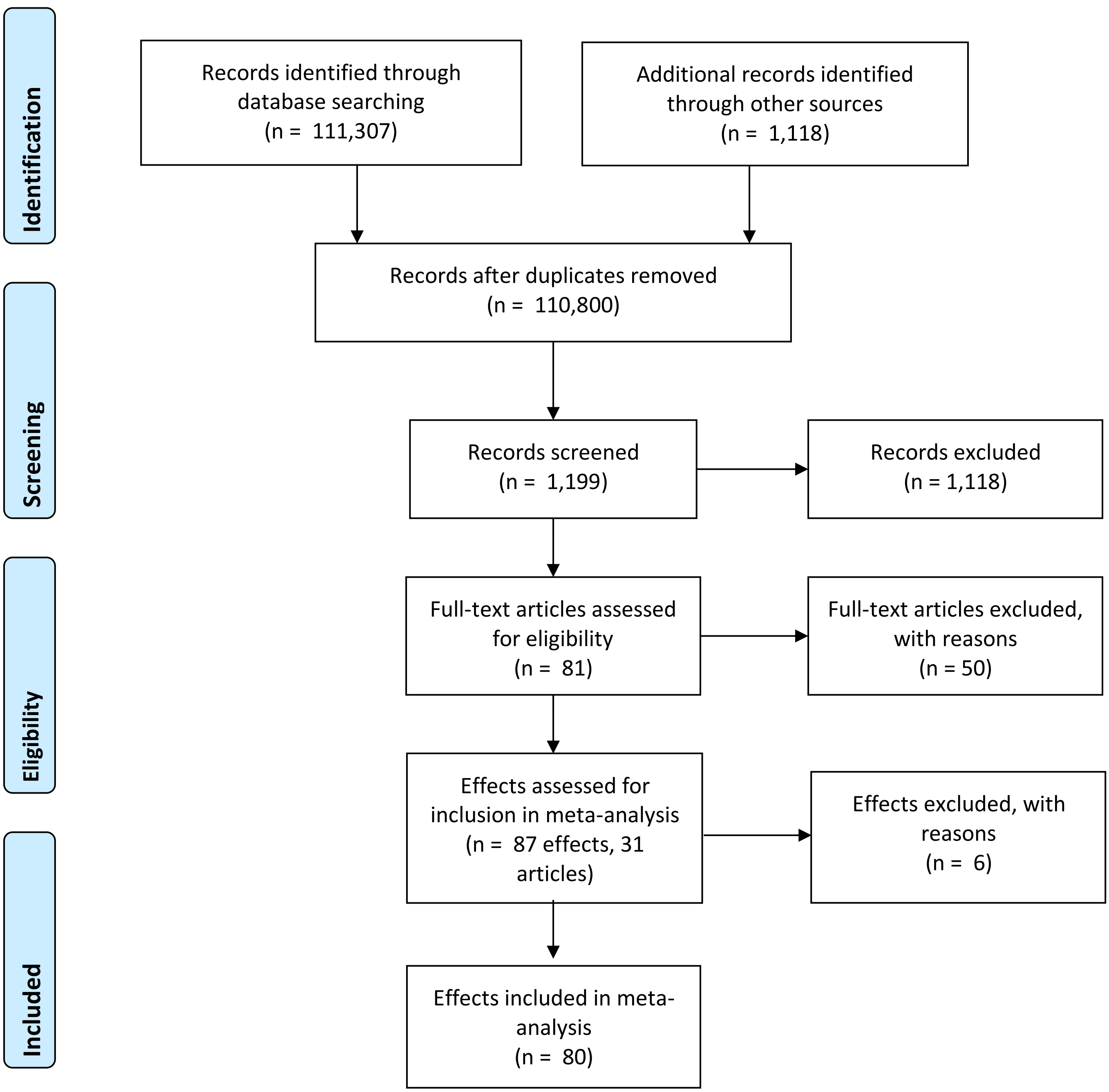


Table 1

Summary of studies included in the meta-analysis. Column headers can be understood as follows: Year (Year of publication), Study Code (Code assigned for analysis, aligns with codes used in figures and elsewhere), Article Name (Authors of the paper), Article DOI (where available), Study (The study/experiment number from the original publication), Within Or Between (whether the design is a between-participant manipulation or within-participant/repeated measures manipulation), Design (whether the study was experimental or quasi-experimental), Independent (whether the measures from a study are independent from one another (i.e., whether multiple measures come from the same participant), where 1 = yes, the measures are independent, and where 0 means they are not independent), Ambient (The nature of the Independent Variable, where 1 = Ambient temperature manipulation, 0 = Haptic temperature manipulation), Prosocial (The nature of the Dependent Variable, where 1 = a prosocial outcome, and 0 = an antisocial outcome), Context (the nature of the social context of the study, where 1 = positive context, 0 = neutral context, -1 = negative context), Result (https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23 links to a file on the OSF containing the raw information used in the calculation of effect sizes – this current view-only link will be replaced by direct links to individual effect sizes prior to publication, but all are currently accessible in this folder), DV (description of the dependent variable), Field (whether the study took place in the field or in lab/lab-like conditions, where 1 = a field study, and 0 = a lab study), N (total number of participants in the study), % Female indicates percentage of participants that were female (a ^ indicates that the values was estimated for individual conditions from the overall %), Effect Size (effect size as *g*), Effect Variance, SE (Standard error of the effect), Moderator (description of potential moderators: what the study context was, and whether the study measured a prosocial or antisocial outcome). A "*" in the results column indicates there was insufficient information about that measure for inclusion in the meta-analysis. For Schneider, Lesko & Garrett (SLG1980_1), there was insufficient variance in responses to include and effect size for this study.

Year	Study Code	ARTICLE NAME	ARTICLE DOI	STUDY	WITHIN OR BETWEEN	DESIGN	Independent	Ambient	Prosocial	Context	RESULT	DV	Field	N	% Female	Effect Size	Effect Variance	SE	Moderator
1980	SLG1980_2	Schneider, Lesko, & Garrett (1980), measure 2	10.1177/0013916580122007	2	between	quasi-experiment, 1x3	1	1	1	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Help, groceries	1	64	-	-0.042	0.175623	0.031	Positive, Prosocial
1980	SLG1980_3	Schneider, Lesko, & Garrett (1980), measure 3	10.1177/0013916580122007	3	between	quasi-experiment, 1x3	1	1	1	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Help, contact lens	1	67	-	0.461	0.157632	0.025	Positive, Prosocial
1980	SLG1980_4	Schneider, Lesko, & Garrett (1980), measure 4	10.1177/0013916580122007	4	between	quasi-experiment, 1x3	1	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Help, questionnaire	1	241	-	-0.21	0.03677	0.001	Neutral, Prosocial
2008	WB2008	Williams & Bargh (2008), S2	10.1126/science.1162548	2	between	experiment, 1x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	1	53	49	0.6907	0.114589	0.013	Neutral, Prosocial
2008	WILLIAMS2008	Williams, Study 6	https://search.proquest.com/openview/715d12dbd345baeea8872b4b52f70376	6	between	experiment, 1x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	1	33	57.6	0.7435	0.196472	0.039	Neutral, Prosocial
2011	KANG2011_1	Kang et al. (2011), S1	10.1093/scan/nsq077	1	between	experiment, 1x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Amount of money invested	0	30	-	1.1427	0.155094	0.024	Neutral, Prosocial
2011	KANG2011_2	Kang et al. (2011), S2	10.1093/scan/nsq077	2	within	experiment, 1x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Amount of money invested	0	23	-	0.0557	0.125048	0.016	Neutral, Prosocial
2011	KGW2012	Kolb, Gockel, Werth (2012), Study 1 a	10.1080/00140139.2012.659763	1	between	experiment, 1x2	0	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Customer oriented helping behaviour	0	69	87	-0.257	0.058462	0.003	Neutral, Prosocial
2011	KGW2012	Kolb, Gockel, Werth (2012), Study 1 b	10.1080/00140139.2012.659763	1	between	experiment, 1x2	0	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Discounts given to customers	0	69	87	-0.55	0.060174	0.004	Neutral, Prosocial
2013	IJ2013_1	IJzerman et al. (2013), Insecure Kids	10.1027/1864-9335/a000142	1	between	experiment, 2x2	1	1	1	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Sharing behaviour (average of stickers and balloons)	0	22	43.3^	-0.479	0.187035	0.035	Negative, Prosocial
2013	IJ2013_2	IJzerman et al. (2013), Secure Kids	10.1027/1864-9335/a000142	1	between	experiment, 2x2	1	1	1	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Sharing behaviour (average of stickers and balloons)	0	38	43.3^	0.642	0.113757	0.013	Positive, Prosocial
2013	SW2013	Storey & Workman (2013)	10.1177/147470491301100106	1	within	experiment, 1x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Cooperation in prisoner's dilemma	0	30	80	0.6104	0.069771	0.005	Neutral, Prosocial
2017	LC2014_3	Lynott, Corker, Connell, O'Brien (2017) - archives (UK)	10.1037/arc0000031	3	between	correlational	0	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	0	305	49	0.1706	0.005003	3E-05	Neutral, Prosocial
2014	LC2014_1	Lynott, Corker, et al. (2014) Kenyon	10.1027/1864-9335/a000187	1	between	experiment, 1x2	0	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	1	306	52.9	-0.265	0.017518	3E-04	Neutral, Prosocial
2014	LC2014_2	Lynott, Corker, et al. (2014) MSU	10.1027/1864-9335/a000187	2	between	experiment, 1x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	1	250	58.7	-0.028	0.020568	4E-04	Neutral, Prosocial
2014	LC2014_3	Lynott, Corker, et al. (2014) UK	10.1027/1864-9335/a000187	3	between	experiment, 1x2	0	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	1	305	51.5	-0.144	0.018839	4E-04	Neutral, Prosocial
2015	L2014	Lynott (2014), unpublished data	NA	1	between	experiment, 1x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	1	113	68.6	0.1819	0.044147	0.002	Neutral, Prosocial
2014	CD2014	Callicoat & Duell (2014)	NA	1	between	experiment, 1x3	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Willingness to help	0	118	83.9	0.1447	0.041774	0.002	Neutral, Prosocial
1972	BARON1972_1	Baron (1972), measure 1a	10.1037/h0032892	1	between	experiment, 2x2	0	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration	0	20	0	0.4232	0.204478	0.042	Negative, Antisocial
1972	BARON1972_2	Baron (1972), measure 1b	10.1037/h0032892	1	between	experiment, 2x2	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration	0	20	0	0.4617	0.205329	0.042	Neutral, Antisocial
1972	BARON1972_1	Baron (1972), measure 2a	10.1037/h0032892	1	between	experiment, 2x2	0	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock intensity	0	20	0	0.2406	0.201447	0.041	Negative, Antisocial
1972	BARON1972_2	Baron (1972), measure 2b	10.1037/h0032892	1	between	experiment, 2x2	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock intensity	0	20	0	0.2406	0.201447	0.041	Neutral, Antisocial
1972	BL1972_1	Baron & Lawton (1972), a	10.3758/BF03335438	1	between	experiment, 2x2	1	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock intensity	0	20	0	-0.841	0.217694	0.047	Negative, Antisocial
1972	BL1972_2	Baron & Lawton (1972), b	10.3758/BF03335438	1	between	experiment, 2x2	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock intensity	0	20	0	0.3761	0.203536	0.041	Neutral, Antisocial
1976	BB1975_1	Baron & Bell (1975) a	10.1037/h0076647	1	between	experiment, 2x2x2	1	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	0.0383	0.250046	0.063	Negative, Antisocial

1976	BB1975_2	Baron & Bell (1975) b	10.1037/h0076647	1	between	experiment, 2x2x2	1	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	0.6517	0.263274	0.069	Negative, Antisocial
1976	BB1975_3	Baron & Bell (1975) c	10.1037/h0076647	1	between	experiment, 2x2x2	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	-0.252	0.251983	0.063	Neutral, Antisocial
1976	BB1975_4	Baron & Bell (1975) d	10.1037/h0076647	1	between	experiment, 2x2x2	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	-0.515	0.258283	0.067	Neutral, Antisocial
1976	BB1976_1	Bell & Baron (1976) a	10.1111/j.1559-1816.1976.tb01308.x	1	between	experiment, 2x2x2	1	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	-1.352	0.307161	0.094	Positive, Antisocial
1976	BB1976_2	Bell & Baron (1976) b	10.1111/j.1559-1816.1976.tb01308.x	1	between	experiment, 2x2x2	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	-0.07	0.250152	0.063	Neutral, Antisocial
1976	BB1976_3	Bell & Baron (1976) c	10.1111/j.1559-1816.1976.tb01308.x	1	between	experiment, 2x2x2	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	-0.153	0.250735	0.063	Neutral, Antisocial
1976	BB1976_4	Bell & Baron (1976) d	10.1111/j.1559-1816.1976.tb01308.x	1	between	experiment, 2x2x2	1	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	0.7808	0.269052	0.072	Negative, Antisocial
1976	BB1976_5	Baron & Bell (1976) Study 1a	10.1037/0022-3514.33.3.245	1	between	experiment, 2x3	1	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	18	48.6	-0.261	0.336165	0.113	Positive, Antisocial
1976	BB1976_6	Baron & Bell (1976) Study 1b	10.1037/0022-3514.33.3.245	1	between	experiment, 2x3	1	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	17	48.6	0.8262	0.397694	0.158	Negative, Antisocial
1976	BB1976_7	Baron & Bell (1976) Study 2a	10.1037/0022-3514.33.3.245	2	between	experiment, 2x2x2	1	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	-0.183	0.251049	0.063	Positive, Antisocial
1976	BB1976_8	Baron & Bell (1976) Study 2b	10.1037/0022-3514.33.3.245	2	between	experiment, 2x2x2	1	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	-0.746	0.267383	0.071	Positive, Antisocial
1976	BB1976_9	Baron & Bell (1976) Study 2c	10.1037/0022-3514.33.3.245	2	between	experiment, 2x2x2	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	0.4841	0.257325	0.066	Neutral, Antisocial
1976	BB1976_10	Baron & Bell (1976) Study 2d	10.1037/0022-3514.33.3.245	2	between	experiment, 2x2x2	1	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	16	0	0.7458	0.267383	0.071	Negative, Antisocial
1976	BARON1976_1	Baron (1976), a1	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Latency to first honk	1	24	0	-0.772	0.179085	0.032	Neutral, Antisocial
1976	BARON1976_2	Baron (1976), a2	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Latency to first honk	1	24	0	-1.137	0.193603	0.037	Neutral, Antisocial
1976	BARON1976_3	Baron (1976), a3	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Latency to first honk	1	24	0	-0.06	0.166741	0.028	Positive, Antisocial
1976	BARON1976_4	Baron (1976), a4	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Latency to first honk	1	24	0	0.2079	0.167568	0.028	Positive, Antisocial
1976	BARON1976_5	Baron (1976), a5	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Latency to first honk	1	24	0	-0.268	0.168159	0.028	Neutral, Antisocial
1976	BARON1976_1	Baron (1976), b1	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Proportion honking	1	24	0	-0.94	0.445583	0.199	Neutral, Antisocial
1976	BARON1976_2	Baron (1976), b2	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Proportion honking	1	24	0	-1.137	0.435813	0.19	Neutral, Antisocial
1976	BARON1976_3	Baron (1976), b3	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Proportion honking	1	24	0	0.4202	0.239311	0.057	Positive, Antisocial
1976	BARON1976_4	Baron (1976), b4	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Proportion honking	1	24	0	0.3822	0.215308	0.046	Positive, Antisocial
1976	BARON1976_5	Baron (1976), b5	10.1111/j.1559-1816.1976.tb01330.x	1	between	quasi-experiment, 2x5	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Proportion honking	1	24	0	0	0.202642	0.041	Neutral, Antisocial
1977	BB1977_1	Bell & Baron (1977), a	10.3758/BF03337050	1	between	experiment, 2x4	1	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	36	0	-0.465	0.114116	0.013	Positive, Antisocial
1977	BB1977_2	Bell & Baron (1977), b	10.3758/BF03337050	1	between	experiment, 2x4	1	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock duration x intensity	0	36	0	0.0202	0.111117	0.012	Negative, Antisocial
1979	PR1979_1	Palmarek & Rule (1979), a	10.1007/BF00994163	1	between	experiment, 2x2	1	1	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Aggressive choice	0	50	0	0.6057	0.236416	0.056	Negative, Antisocial

1979	PR1979_2	Palmarek & Rule (1979), b	10.1007/BF00994163	1	between	experiment, 2x2	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Aggressive choice	0	50	0	-1.073	0.286594	0.082	Neutral, Antisocial
1980	BELL_1	Bell (1980), a	10.1080/00224545.1980.9924227	1	between	experiment, 2x2x2	1	1	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Retaliation against experimenter	0	40	0	1.1082	0.115352	0.013	Positive, Antisocial
1980	BELL_2	Bell (1980), b	10.1080/00224545.1980.9924227	1	between	experiment, 2x2x2	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Retaliation against experimenter	0	40	0	-0.334	0.101391	0.01	Neutral, Antisocial
1986	KM1986	Kenrick & McFarlane (1986)	10.1177/0013916586182002	1	between	correlational	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Aggregate of latency to honk and number of honks	1	75	48	-0.73	0.04961	0.002	Neutral, Antisocial
1994	VV1994	Vrij & Van der Steen (1994), measure 3	10.1002/casp.2450040505	1	between	experiment, 1x2	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shooting behaviour (shot vs. not)	0	38	21	-0.333	0.133146	0.018	Neutral, Antisocial
1994	VV1994	Vrij & Van der Steen (1994), measure 1	10.1002/casp.2450040505	1	between	experiment, 1x2	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Tendency to shoot (self-report)	0	38	21	-0.78	0.11	0.012	Neutral, Antisocial
1994	VV1994	Vrij & Van der Steen (1994), measure 2	10.1002/casp.2450040505	1	between	experiment, 1x2	0	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Firearm position (holstered vs. not)	0	38	21	0.7912	0.192028	0.037	Neutral, Antisocial
2012	B2012_1	Breines (2012), Study 3a1	https://escholarship.org/uc/item/3sw4v2f9	3	between	experiment, 2x3	1	0	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Punishment severity	0	55	58^	-0.499	0.113003	0.013	Positive, Antisocial
2012	B2012_2	Breines (2012), Study 3a2	https://escholarship.org/uc/item/3sw4v2f9	3	between	experiment, 2x3	1	0	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Punishment severity	0	58	58^	-0.18	0.099428	0.01	Negative, Antisocial
2014	FM2014_1	Fay & Maner (2014), S2a No rejection	10.1016/j.jesp.2013.10.006	2	between	quasi-experiment, 2x2xFNE	1	0	0	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Noise blast (volume and duration composite)	0	54	82.2	0.4635	0.08774	0.008	Positive, Antisocial
2014	FM2014_2	Fay & Maner (2014), S2b Rejection	10.1016/j.jesp.2013.10.006	2	between	quasi-experiment, 2x2xFNE	1	0	0	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Noise blast (volume and duration composite)	0	53	82.2	-0.5	0.069426	0.005	Negative, Antisocial
2014	GKW_2014	Gockel, Kolb & Werth (2014) a	10.1371/journal.pone.0096231	1	between	experiment, 1x3	1	1	0	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Degree of penalty	0	133	79	0.6025	0.047517	0.002	Neutral, Antisocial
1994	VGFKV1994_1	van Goozen, Frijda, Kindt,and van de Poll (1994) a measure 2	10.1002/1098-2337(1994)20:2<79::AID-AB2480200202>3.0.CO;2-K	1	between	experiment, 2x2	0	1	1	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Percent who gave	0	30	100	-0.176	0.194813	0.038	Positive, Prosocial
1994	VGFKV1994_2	van Goozen, Frijda, Kindt,and van de Poll (1994) b measure 2	10.1002/1098-2337(1994)20:2<79::AID-AB2480200202>3.0.CO;2-K	1	between	experiment, 2x2	0	1	1	-1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Percent who gave	0	30	100	0	0.350727	0.123	Negative, Prosocial
2014	FAY2014_1	Fay (2014), Study 3a	http://purl.flvc.org/fsu/fd/FSU_migr_etd-8778	3	between	quasi-experiment, 2x2xFNE	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Money awarded	0	49	27.1^	0.0727	0.083387	0.007	Neutral, Prosocial
2014	FAY2014_2	Fay (2014), Study 3b	http://purl.flvc.org/fsu/fd/FSU_migr_etd-8778	3	between	quasi-experiment, 2x2xFNE	1	0	1	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Money awarded	0	43	27.1^	-0.15	0.097548	0.01	Positive, Prosocial
2014	FAY2014_3	Fay (2014), Study 5a	http://purl.flvc.org/fsu/fd/FSU_migr_etd-8778	5	between	quasi-experiment, 2x2xFNE	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Money awarded	0	108	27.1^	0.3087	0.037235	0.001	Neutral, Prosocial
2014	FAY2014_4	Fay (2014), Study 5b	http://purl.flvc.org/fsu/fd/FSU_migr_etd-8778	5	between	quasi-experiment, 2x2xFNE	1	0	1	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Money awarded	0	110	27.1^	-0.134	0.03744	0.001	Positive, Prosocial
2015	STE2015_1	Steidle and Werth (2012)	NA	1	between	experiment, 2x2	0	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Percent helping	0	148	-	-0.202	0.045859	0.002	Neutral, Prosocial
2015	STE2015_1	Steidle and Werth (2012)	NA	1	between	experiment, 2x2	0	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Minutes helping	0	148	-	0.0287	0.027987	8E-04	Neutral, Prosocial
2017	LC2014_1	Lynott, Corker, Connell, O'Brien (2017) - archives (Kenyon)	10.1037/arc0000031	1	between	correlational	0	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	1	306	49	0.0425	0.007014	5E-05	Neutral, Prosocial
2017	MM2018_1	Miyajima & Meng (2018) a female participants	10.1186/s13104-017-2972-3	1	between	experiment, 2x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Helping behaviour	0	38	100	0.6564	0.114646	0.013	Neutral, Prosocial
2017	MM2018_2	Miyajima & Meng (2018) b male participants	10.1186/s13104-017-2972-3	1	between	experiment, 2x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Helping behaviour	0	31	0	-0.192	0.133945	0.018	Neutral, Prosocial
2015	WHvE2015_1	Willemse, Heylan & van Erp (2015) s1a control	10.1109/ACII.2015.7344547	1	between	experiment, 2x3	1	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Prosocial behaviour (prisoner's dilemma)	0	25	32.9^	0.3859	0.191418	0.037	Neutral, Prosocial
2015	WHvE2015_2	Willemse, Heylan & van Erp (2015) s1b artificial heat	10.1109/ACII.2015.7344547	1	between	experiment, 2x3	1	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Prosocial behaviour (prisoner's dilemma)	0	28	32.9^	-0.308	0.175511	0.031	Neutral, Prosocial
2015	WHvE2015_3	Willemse, Heylan & van Erp (2015) s1c body heat	10.1109/ACII.2015.7344547	1	between	experiment, 2x3	1	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Prosocial behaviour (prisoner's dilemma)	0	32	32.9^	0.1189	0.152875	0.023	Neutral, Prosocial

2017	BK2017	Belkin & Kouchaki (2017) S3a	10.1002/ejsp.2242	3	between	quasi-experiment, 1x2	1	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Helping (Number of questions answered)	1	160	32.9^	-1.232	0.065193	0.004	Neutral, Prosocial
2017	RLY2017	Rai, Lin & Yang (2017) S3a	10.1108/JCM-07-2015-1505	3	between	experiment, 1x2	1	1	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Money donated	0	106	72.6	-0.486	0.038849	0.002	Neutral, Prosocial
2018	CHAB2019	Chabris et al. (2019), S2a	10.1027/1864-9335/a000361	2	between	experiment, 1x2	1	0	1	0	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Choice of gift for friend	1	168	54	0.0477	0.029089	8E-04	Neutral, Prosocial
1980	SLG1980_1	Schneider, Lesko, & Garrett (1980), measure 1	10.1177/0013916580122007	1	between	quasi-experiment, 1x3	1	1	1	1	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Help, crutches	1	68	-	-	-	-	Positive, Prosocial
1999	B1999_1	Boyanowsky (1999), Study 2a		2	between	experiment, 3x2	1	1	0	-	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock intensity*	0	30	-	-	-	-	Neutral, Antisocial
1999	B1999_2	Boyanowsky (1999), Study 2b		2	between	experiment, 3x2	1	1	0	-	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock intensity*	0	30	-	-	-	-	Neutral, Antisocial
1999	B1999_3	Boyanowsky (1999), Study 3a		2	between	experiment, 2x3	1	1	0	-	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock intensity*	0	30	-	-	-	-	Neutral, Antisocial
1999	B1999_3	Boyanowsky (1999), Study 3b		2	between	experiment, 2x3	1	1	0	-	https://osf.io/ghym5/?view_only=8b3288219cbb487397a0fdd1c3260d23	Shock intensity*	0	30	-	-	-	-	Neutral, Antisocial

Table 2

Effect Size Estimates from Moderation Analysis (Type of Dependent Variable × Social Context)

	Antisocial DV	Prosocial DV
Positive Context	-0.05 [-0.39, 0.29] *k* = 12	0.09 [-0.32, 0.50] *k* = 2
Neutral Context	-0.23 [-0.50, 0.04] *k* = 21	0.06 [-0.13, 0.25] *k* = 27
Negative Context	0.14 [-0.20, 0.47] *k* = 12	-0.29 [-1.16, 0.59] *k* = 6

Note. Estimates (Hedge's *g*) are from multivariate model. DV = dependent variable.

Table 3

PET-PEESE Meta-Regression and Funnel Plot Asymmetry Test Results

	PET		PEESE	
	Intercept	FAT	Intercept	FAT
Overall	-0.080	0.218	-0.051	0.305
	[-0.29, 0.13]	[-0.56, 0.99]	[-0.19, 0.09]	[-1.01, 1.62]
Prosocial	-0.157	0.738	-0.088	1.538
	[-0.41, 0.10]	[-0.46, 1.94]	[-0.25, 0.07]	[-0.80, 3.88]
Antisocial	-0.011	-0.133	-0.027	-0.220
	[-0.70, 0.68]	[-1.94, 1.67]	[-0.42, 0.37]	[-2.56, 2.12]

Note. FAT = Funnel plot Asymmetry Test, PET = Precision Effect Test, PEESE = Precision Effect Estimate with Standard Error. Values are weighted least squares regression coefficients [95% confidence intervals].

Table 4

Summary of findings table indicating overall level of confidence in findings (Certainty of Evidence) for each group of studies. ***High certainty*** *indicates "We are very confident that the true effect lies close to that of the estimate of the effect",* ***Moderate certainty*** *indicates "We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different",* ***Low certainty*** *indicates "Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect", and* ***Very Low certainty*** *indicates "We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect". Superscript notes indicate contributors to estimates of certainty: a. Common for studies to be without pre-registered analysis plans, b. Generally small sample sizes, c. Differences in populations, d. Differences in outcomes, e. Differences in interventions (i.e., level of temperature manipulations), f. Considerable variability in effect sizes*

Certainty assessment						№ of patients		Effect	Certainty
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	hot	cold	Absolute (95% CI)	
Prosocial outcomes, Haptic (assessed with: Behavioural choice e.g., gift for friend)									
17	Random allocation, majority between-participant designs	serious [a]	not serious	not serious	not serious	876	888	**0.009 Hedge's g lower** (0.205 lower to 0.187 higher)	⊕⊕⊕◯ MODERATE
Prosocial outcomes, Ambient (assessed with: Behavioural Choice e.g., gift for friend)									
18	Random allocation, all between-participant designs	serious [a,b]	serious [c,d,e]	serious [c,d,e]	serious [b]	528	478	**0.179 Hedge's g lower** (0.437 lower to 0.078 higher)	⊕◯◯◯ VERY LOW
Antisocial outcomes, Haptic (assessed with: Behavioural Choice, e.g., severity of punishment)									
4	Random allocation, all between-participant designs	serious [a,b]	serious [f]	not serious	serious [e]	81	107	**0.178 Hedge's g lower** (0.672 lower to 0.316 higher)	⊕◯◯◯ VERY LOW
Antisocial outcomes, Ambient (assessed with: Behavioural choice, e.g., severity of punishment)									

Certainty assessment						№ of patients		Effect	Certainty
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	hot	cold	Absolute (95% CI)	
41	Random allocation, all between-participant designs. Some quasi-experimental studies	very serious [a,b,e]	serious [a,b,f]	not serious	serious [f]	482	532	**0.061 Hedge's g lower** (0.26 lower to 0.138 higher)	⊕◯◯◯ VERY LOW

Figure 1

Meta-analytic estimates of effects (Hedge's *g*) from multivariate moderation analysis (Type of Dependent Variable × Social Context).

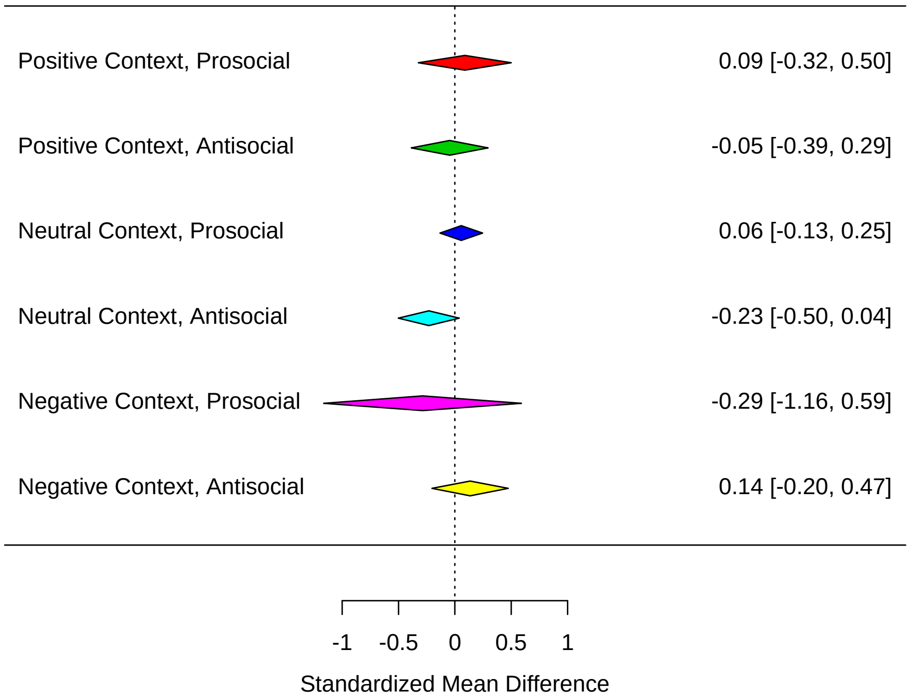
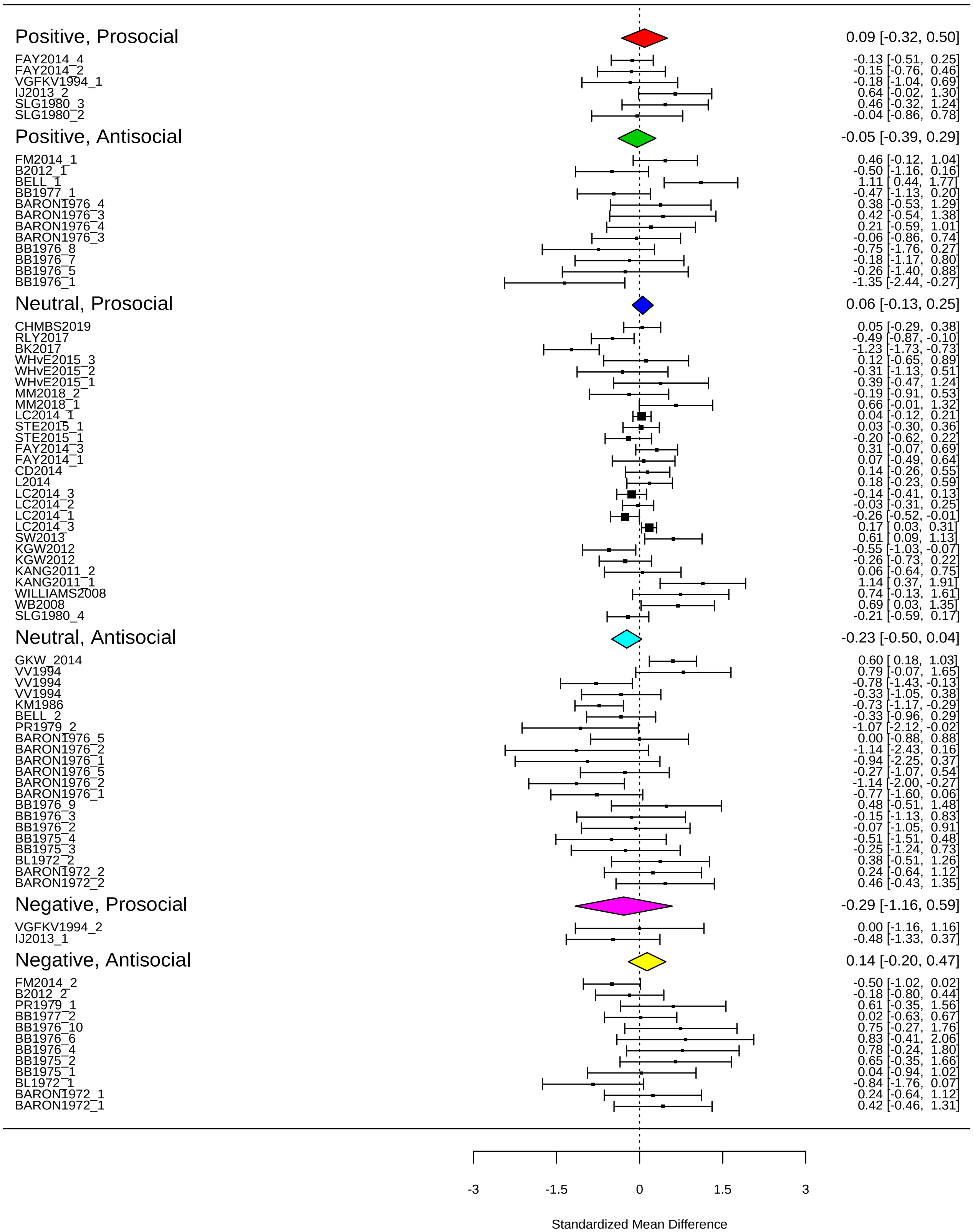


Figure 2

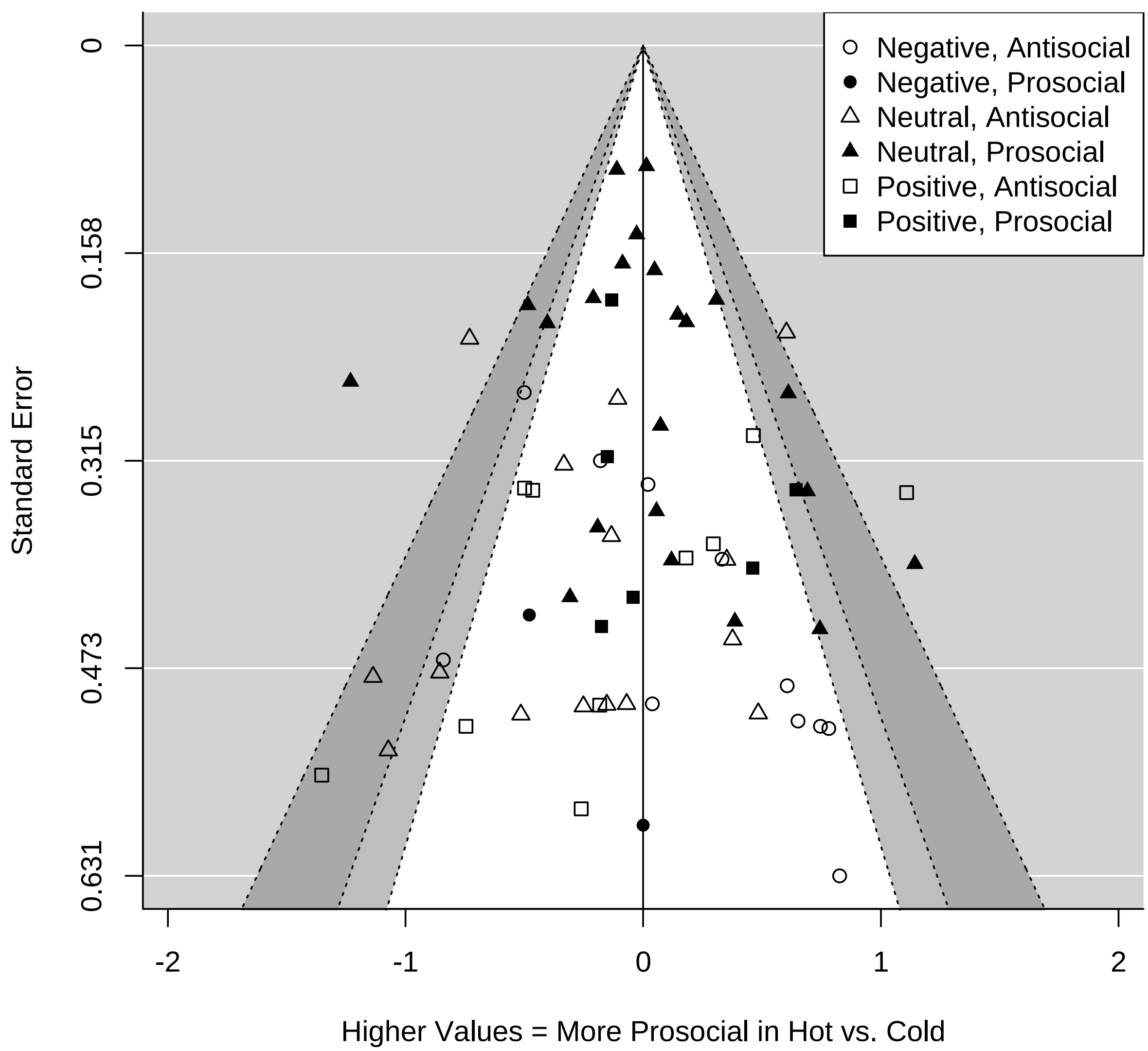
Meta-analytic estimates of effects (Hedge's *g*) from multivariate moderation analysis (Type of Dependent Variable × Social Context).



Positive, Prosocial 0.09 [-0.32, 0.50]
FAY2014_4 -0.13 [-0.51, 0.25]
FAY2014_2 -0.15 [-0.76, 0.46]
VGFKV1994_1 -0.18 [-1.04, 0.69]
IJ2013_2 0.64 [-0.02, 1.30]
SLG1980_3 0.46 [-0.32, 1.24]
SLG1980_2 -0.04 [-0.86, 0.78]
Positive, Antisocial -0.05 [-0.39, 0.29]
FM2014_1 0.46 [-0.12, 1.04]
B2012_1 -0.50 [-1.16, 0.16]
BELL_1 1.11 [0.44, 1.77]
BB1977_1 -0.47 [-1.13, 0.20]
BARON1976_4 0.38 [-0.53, 1.29]
BARON1976_3 0.42 [-0.54, 1.38]
BARON1976_4 0.21 [-0.59, 1.01]
BARON1976_3 -0.06 [-0.86, 0.74]
BB1976_8 -0.75 [-1.76, 0.27]
BB1976_7 -0.18 [-1.17, 0.80]
BB1976_5 -0.26 [-1.40, 0.88]
BB1976_1 -1.35 [-2.44, -0.27]
Neutral, Prosocial 0.06 [-0.13, 0.25]
CHMBS2019 0.05 [-0.29, 0.38]
RLY2017 -0.49 [-0.87, -0.10]
BK2017 -1.23 [-1.73, -0.73]
WHvE2015_3 0.12 [-0.65, 0.89]
WHvE2015_2 -0.31 [-1.13, 0.51]
WHvE2015_1 0.39 [-0.47, 1.24]
MM2018_2 -0.19 [-0.91, 0.53]
MM2018_1 0.66 [-0.01, 1.32]
LC2014_1 0.04 [-0.12, 0.21]
STE2015_1 0.03 [-0.30, 0.36]
STE2015_1 -0.20 [-0.62, 0.22]
FAY2014_3 0.31 [-0.07, 0.69]
FAY2014_1 0.07 [-0.49, 0.64]
CD2014 0.14 [-0.26, 0.55]
L2014 0.18 [-0.23, 0.59]
LC2014_3 -0.14 [-0.41, 0.13]
LC2014_2 -0.03 [-0.31, 0.25]
LC2014_1 -0.26 [-0.52, -0.01]
LC2014_3 0.17 [0.03, 0.31]
SW2013 0.61 [0.09, 1.13]
KGW2012 -0.55 [-1.03, -0.07]
KGW2012 -0.26 [-0.73, 0.22]
KANG2011_2 0.06 [-0.64, 0.75]
KANG2011_1 1.14 [0.37, 1.91]
WILLIAMS2008 0.74 [-0.13, 1.61]
WB2008 0.69 [0.03, 1.35]
SLG1980_4 -0.21 [-0.59, 0.17]
Neutral, Antisocial -0.23 [-0.50, 0.04]
GKW_2014 0.60 [0.18, 1.03]
VV1994 0.79 [-0.07, 1.65]
VV1994 -0.78 [-1.43, -0.13]
VV1994 -0.33 [-1.05, 0.38]
KM1986 -0.73 [-1.17, -0.29]
BELL_2 -0.33 [-0.96, 0.29]
PR1979_2 -1.07 [-2.12, -0.02]
BARON1976_5 0.00 [-0.88, 0.88]
BARON1976_2 -1.14 [-2.43, 0.16]
BARON1976_1 -0.94 [-2.25, 0.37]
BARON1976_5 -0.27 [-1.07, 0.54]
BARON1976_2 -1.14 [-2.00, -0.27]
BARON1976_1 -0.77 [-1.60, 0.06]
BB1976_9 0.48 [-0.51, 1.48]
BB1976_3 -0.15 [-1.13, 0.83]
BB1976_2 -0.07 [-1.05, 0.91]
BB1975_4 -0.51 [-1.51, 0.48]
BB1975_3 -0.25 [-1.24, 0.73]
BL1972_2 0.38 [-0.51, 1.26]
BARON1972_2 0.24 [-0.64, 1.12]
BARON1972_2 0.46 [-0.43, 1.35]
Negative, Prosocial -0.29 [-1.16, 0.59]
VGFKV1994_2 0.00 [-1.16, 1.16]
IJ2013_1 -0.48 [-1.33, 0.37]
Negative, Antisocial 0.14 [-0.20, 0.47]
FM2014_2 -0.50 [-1.02, 0.02]
B2012_2 -0.18 [-0.80, 0.44]
PR1979_1 0.61 [-0.35, 1.56]
BB1977_2 0.02 [-0.63, 0.67]
BB1976_10 0.75 [-0.27, 1.76]
BB1976_6 0.83 [-0.41, 2.06]
BB1976_4 0.78 [-0.24, 1.80]
BB1975_2 0.65 [-0.35, 1.66]
BB1975_1 0.04 [-0.94, 1.02]
BL1972_1 -0.84 [-1.76, 0.07]
BARON1972_1 0.24 [-0.64, 1.12]
BARON1972_1 0.42 [-0.46, 1.31]
-3
-1.5
0
1.5
3
Standardized Mean Difference

Figure 3

A contour-enhanced funnel plot of effect sizes (x-axis) vs. their standard errors (a measure of precision). All studies coded such that higher values mean more prosocial behaviour in hot vs. cold conditions. Plot represents 67 independent data points (dependent effect sizes within studies are averaged together). Moving from the centre, the light grey zones show effects between p = .1 and p = .05, and the dark grey zones show effects between p = .05 and p = .01. Asymmetry in funnel plots can be a sign of publication bias because asymmetry occurs when more precise studies estimate effects closer to zero. Recall that each point represents a comparison between two conditions (hot vs. cold), but within studies, researchers often used a 2 × 2 or more complicated design.



Negative, Antisocial
Negative, Prosocial
Neutral, Antisocial
Neutral, Prosocial
Positive, Antisocial
Positive, Prosocial
Standard Error
0
0.158
0.315
0.473
0.631
-2
-1
0
1
2
Higher Values = More Prosocial in Hot vs. Cold

Figure 4

Risk of bias across studies grouped by domain from the RoB 2 tool (Sterne et al., 2019). Darker greys indicate a greater prominence of risk