Check for updates

RESEARCH ARTICLE

# Evidence from over 30 studies suggests the existence of evolved adaptations to war in humans

[version 1; peer review: 1 approved, 1 approved with reservations, 1 not approved]

Henrikas Bartusevičius, Samira Aminihajibashi, Stefan Goetz, Eric Skoog, Thomas Hagen

Peace Research Institute Oslo, Oslo, Oslo, Norway

**v1** **First published:** 19 Mar 2026, **6**:78
https://doi.org/10.12688/openreseurope.21936.1
**Latest published:** 19 Mar 2026, **6**:78
https://doi.org/10.12688/openreseurope.21936.1

## Abstract

### Background

Was small-scale war a recurrent feature of life during human evolution, and—consequently—have humans adapted to war? This question has generated major scholarly debate, primarily informed by archaeology, ethnography, and primatology. We examined adaptation to war in contemporary humans via cognitive psychology experiments. If war recurred during evolution, then our minds are likely equipped with efficient, complex, and specialized mechanisms for effective navigation of intergroup coalitional aggression.

### Method

Fourteen hypotheses were derived positing the existence of rapid and accurate coalition detection, enumeration, and formidability assessment mechanisms. Twenty-five exploratory experiments with ca. 5000 US residents custom-fitted established attentional bias, enumeration, and rating paradigms for large-scale online administration, and preliminary tested the hypotheses. Seven preregistered studies with probability samples totaling ca. 12000 US residents tested the hypotheses in a confirmatory fashion.

### Results

The confirmatory studies supported 11 of the 14 hypotheses. Participants automatically attended to, efficiently enumerated, and made nuanced assessments of rapidly presented schematic coalitions. Male participants enumerated the coalitions more rapidly and

**Open Peer Review**

**Approval Status** ✗ ✓ ?

| | 1 | 2 | 3 |
|---|---|---|---|
| **version 1** 19 Mar 2026 | ✗ view | ✓ view | ? view |

1. **Anthony Lopez**, Washington State University, Pullman, USA
2. **Aaron Sell**, Heidelberg University, Tiffin, USA
3. **Gijs Schumacher**, University of Amsterdam, Amsterdam, The Netherlands

Any reports and responses or comments on the article can be found at the end of the article.

accurately, female participants perceived them as more formidable, and female and less formidable participants overestimated their numerical size.

**Conclusions**

Combined with multiple tests of alternative explanations, the evidence suggests the existence of specialized adaptations to intergroup coalitional aggression, consistent with the broader hypothesis that human prehistory was characterized by war.

**Keywords**
war; coalitional aggression; human evolution; evolved adaptations; psychological mechanisms; formidability; fighting capacity; strength; enumeration; automatic attention; cognitive psychology experiments; prehistory; Pleistocene

This article is included in the European Research Council (ERC) gateway.

This article is included in the Horizon 2020 gateway.

**Corresponding author:** Henrikas Bartusevičius (henrikas@prio.org)

**Author roles: Bartusevičius H**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Aminihajibashi S**: Conceptualization, Investigation, Methodology, Writing – Review & Editing; **Goetz S**: Conceptualization, Investigation, Methodology, Writing – Review & Editing; **Skoog E**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Hagen T**: Data Curation, Investigation, Methodology, Resources, Software, Validation

**Competing interests:** No competing interests were disclosed.

**Grant information:** This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 950313)
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2026 Bartusevičius H *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Bartusevičius H, Aminihajibashi S, Goetz S *et al.* **Evidence from over 30 studies suggests the existence of evolved adaptations to war in humans [version 1; peer review: 1 approved, 1 approved with reservations, 1 not approved]** Open Research Europe 2026, **6**:78 https://doi.org/10.12688/openreseurope.21936.1

**First published:** 19 Mar 2026, **6**:78 https://doi.org/10.12688/openreseurope.21936.1

> *Artifacts tell us that violent conflict appeared with the very first man … war grew out of the same base instinct for domination or conquest that had caused conflicts among the simplest tribes, an old pattern amplified by new capabilities and without new constraints*—Barack Obama, Speech at the Hiroshima Peace Memorial (The New York Times, 2016).

## Introduction

The question of whether war characterized human evolution has generated major interdisciplinary debate, involving anthropologists (e.g., Boehm, 2012; Fry & Söderberg, 2013), archaeologists (e.g., Keeley, 1996; Kelly, 2005), economists (e.g., Bowles, 2009; Choi & Bowles, 2007), historians (e.g., Gat, 2008), primatologists (e.g., Wilson *et al.*, 2014; Wilson & Wrangham, 2003), psychologists (e.g., Pinker, 2011a; Sell *et al.*, 2009), and political scientists (e.g., Fukuyama, 1998; Goldstein, 2011). It has also shaped public understandings of human nature (e.g., Diab, 2008; Glowacki, 2016; Gorman, 2016; Horgan, 2016; Johnson, 2014; Pinker, 2011b; Romm, 2016) and influenced political and military thought. Barack Obama referred to the war instinct not only at the Hiroshima Peace Memorial, but also while accepting the Nobel Prize (Obama, 2009). A report for the Pentagon on the "root causes of war" examined archaeological and anthropological literature to argue that "warfare is inherent to human species" (Haas & Piscitelli, 2015: 168).

However, the scholarly debate remains unsettled. Some scholars claim that humans "come from a long evolutionary history of war" (Sell *et al.*, 2017: 334) and are therefore equipped with adaptations to war (see also Bowles, 2009; Choi & Bowles, 2007; Gat, 2008; Goldstein, 2011; Keeley, 1996; LeBlanc & Register, 2003; Manson *et al.*, 1991; Pinker, 2011a; Sell *et al.*, 2009; Wilson & Wrangham, 2003; Wrangham, 1999; Wrangham & Peterson, 1996), and others argue that war is a "recent innovation in human history" (Chomsky, 2015) and that we therefore could not have evolved adaptations to war (see also Fry, 2006; Fry, 2007; Fry, 2015; Fry & Söderberg, 2013; Fuentes, 2017; Kelly, 2000; Kelly, 2005). "War" is defined here in a generic sense, as "events in which coalitions of members of a group seek to inflict bodily harm on one or more members of another group" (Bowles, 2009: 1294).

The debate essentially revolves around the question of whether war was a recurrent feature of life in the Pleistocene, and potentially earlier (up to 6–8 million years ago, when the hominin lineage diverged from that of the chimpanzee and the bonobo). The relevant archaeological evidence falls into several categories: representations of war in rock and cave art, technology (e.g., intra-specific killing-specific tools, such as combat maces), settlement data (e.g., fortifications), and skeletal remains (Ferguson, 2015: 193–196; Haas & Piscitelli, 2015: 176; Pinker, 2011a: 58). The latter category includes evidence of violent trauma or death, including projectiles embedded in bones, parry fractures of the ulna, cranial compression fractures, dismemberment, defleshing, and multiple burials (Keeley, 1996; LeBlanc, 1999; LeBlanc & Register, 2003; Martin & Harrod, 2015; Vencl, 1984; Walker, 2001). Archaeologists agree that the evidence is sufficient to infer the existence of war from 8000 BC onwards (Haas & Piscitelli, 2015: 177–178). This date broadly overlaps with the beginning of the Holocene, marking the gradual transition from a nomadic lifestyle of foraging to sedentary lifestyle of agriculture. Since most of the evolution of the genus Homo occurred in the Pleistocene and since the Holocene represents too short a period for the evolution of complex traits (Cosmides *et al.*, 1992: 5), it is commonly argued that the archaeological evidence relevant for inferences about ancestral war must come from at least the Pleistocene (Haas & Piscitelli, 2015: 174–5). However, archaeological evidence from sites dated before 8,000 BC is both patchier and more ambiguous (Haas & Piscitelli, 2015). The distinctive signs of war, such as fortifications (Pinker, 2011a: 58), do not appear until sedentism in the late Pleistocene or early Holocene, and the interpretations of the Pleistocene skeletal remains diverge (Lahr *et al.*, 2016; Stojanowski *et al.*, 2016). A limitation of skeletal data is that they hardly allow distinguishing between trauma due to coalitional aggression vs dyadic fights. In many cases, we cannot even infer whether trauma comes from intra-specific violence or alternative causes, such as hunting accidents or burial rituals (Martin & Harrod, 2015; Walker, 2001: 574).

Given the limitations in archaeological evidence, scholars often turn to the ethnographic record to make inferences about ancestral war (Bowles, 2006; Bowles, 2009; Guilaine & Zammit, 2005; Keeley, 1996; LeBlanc, 1999; Otterbein, 1970; Otterbein, 2004; Pinker, 2011a). This record includes small-scale societies in which the lifestyle presumably mirrors that of ancestral hominins. If these societies wage war, then our Pleistocene ancestors probably also did so. Akin to archaeology, however, little agreement exists as to whether small-scale societies are warring or not, from reports claiming that they "typically have had fairly frequent warfare" (Ember, 1978: 447) to those arguing that they are "typically non-warring" (Fry, 2007: 143) (see also Boehm, 2015; Bowles, 2006; Bowles, 2009; Fry & Söderberg, 2013; Keeley, 1996; Kelly, 2000; Kelly, 2005; Otterbein, 1970; Otterbein, 2004; Pinker, 2011a; Wrangham & Glowacki, 2012). These debates involve multiple issues, including what constitutes war and which societies are appropriate to study for gaining insights about ancestral war (Wrangham & Glowacki, 2012: 11–12; see also Boehm, 2015; Fry & Söderberg, 2013). A limitation of ethnographic evidence is that it cannot tell the extent to which the historically known societies resemble ancestral hominins: "[h]unters and gatherers of the 'ethnographic present' may be profoundly different from hunters and gatherers of the more distant archaeological past", because they have been directly or indirectly touched by either early states or colonialists (Haas & Piscitelli, 2015: 174–175; see also Wobst, 1978). The few (recently) existing hunter-gatherers (e.g., the Kalahari !Kung; see Shostak, 1981) who did manage to escape the stronger influences of state or agricultural societies potentially did so because they lived in harsh environments, which may poorly represent those in which Pleistocene hominins lived (Marlowe, 2005).

Scholars also use observations of other primates, most notably intergroup coalitional aggression among chimpanzees (e.g., Goodall, 1986; Wilson *et al.*, 2014), to draw inferences about

human war (Manson *et al.*, 1991; Wilson & Wrangham, 2003; Wrangham, 1999; Wrangham & Glowacki, 2012; Wrangham & Peterson, 1996). At least two arguments underlie this comparison. The first is phylogenetic, suggesting that human war builds on "pre-human tendencies" (Wrangham, 1999: 2). Despite the strong pressure for the evolution of coalitional aggression among many animal species, it is an extremely rare behavioral trait (Tooby & Cosmides, 2025). Given this, and invoking the parsimony principle, intergroup coalitional aggression more likely appeared in the human-chimpanzee ancestor rather than independently in each lineage. If true, then human war builds on at least 6 million years old pre-adaptation, characterizing the coalitional aggression observed among contemporary chimpanzees. The second argument is behavioral-ecological, suggesting that human war evolved in response to the same constellation of factors that promoted the evolution of chimpanzee war (Kelly, 2005: 1): both chimpanzee and human males (in small-scale societies) appear to compete over territory, food, and females and show a preference for low-risk attacks and intergroup dominance (Wilson & Wrangham, 2003: 363). Both arguments have faced challenges. The phylogenetic argument faces the absence of bonobo war. The conservative reconstruction principle would require us to conclude that the human-chimpanzee ancestor "was only as prone to intergroup conflict as is seen at the bonobo level, which would preclude patrols and the surprise-attack, lethal raiding type of 'warfare' practiced by chimpanzees and some human foragers" (Boehm, 2015: 317). The behavioral-ecological argument faces objections that point to differences in ecologies and evolutionary trajectories of humans and chimpanzees (for a review, see Ferguson, 2011). In a *Science* special issue devoted to *Ardipithecus ramidus*, scholars claim that the last common ancestor of humans and chimpanzees was not chimpanzee-like and that both hominins and extant African apes are each highly specialized and through different evolutionary pathways" (White *et al.*, 2009: 5). Notably, "the excessively aggressive intermale and intergroup behavior seen in modern P. troglodytes is unique to that lineage and … this derived condition compromises the living chimpanzee as a behavioral model for the ancestral [hominin] condition" (Suwa *et al.*, 2009: 57).

In the present research, we propose to directly probe the existence of adaptations to small-scale war in contemporary humans via cognitive psychology experiments. If we come from a long evolutionary history of war, then our minds are likely equipped with evolved machinery enabling the effective navigation of war: efficient, complex, and specialized mechanisms promoting planning, execution, and defense against intergroup coalitional aggression.

Phylogenetic evidence suggests that human evolution has been characterized by dyadic or one-on-one aggression (Gomez *et al.*, 2016). In ancestral one-on-one conflicts, assessing the formidability of opponents before potential combat was vital: fighting against stronger opponents could result in injury and death, while deferring to weaker ones could forego resources and status. Selection thus likely favored mechanisms for extracting formidability information from evolutionarily relevant cues (Sell *et al.*, 2009). Research shows that people can predict upper-body strength of men from static images of their faces and independently of their height, weight, and age (Sell *et al.*, 2009; see also Fink *et al.*, 2007). Ratings of upper-body strength strongly correlate with ratings of fighting capacity (Sell *et al.*, 2009), suggesting that the former serves as a key visual cue of the latter. Based on face photographs, people can also predict the fighting ability of men as rated by their acquaintances. People can even assess strength from voice (Sell *et al.*, 2010). In turn, adaptive formidability assessment mechanisms should track evolutionarily relevant variables to moderate the assessments. Studies show that individual formidability, social support, the presence of allies, and many other related variables reduce envisioned formidability of hypothetical adversaries (Fessler & Holbrook, 2013a; Fessler & Holbrook, 2013b; Fessler & Holbrook, 2014; Fessler & Holbrook, 2016; Fessler *et al.*, 2016; Fessler *et al.*, 2014; Fessler *et al.*, 2012; Fessler *et al.*, 2014a; Fessler *et al.*, 2014b; Holbrook *et al.*, 2016a; Holbrook *et al.*, 2016b; Holbrook *et al.*, 2017). Many of these findings have been replicated across diverse samples, including small-scale societies (Sell *et al.*, 2009; Sell *et al.*, 2010; von Rueden *et al.*, 2008).

Here, we examine whether humans are equipped with analogous adaptations in the domain of coalitional, not interpersonal, aggression. We focus, specifically, on coalition detection, enumeration, and formidability assessment mechanisms, which likely helped ancestral hominins to avoid disadvantageous, and initiate advantageous, intergroup fights. These mechanisms, if revealed, may constitute signature evidence that human evolution has been shaped by war.

Some findings already hint at some such mechanisms. For example, physically stronger men are more likely to believe that the use of force in modern inter-state armed conflicts will be successful (Sell *et al.*, 2009; see also Bartusevičius, 2021; Price *et al.*, 2012; Sell *et al.*, 2017). The strength of one individual is virtually inconsequential for conflicts involving modern large-scale armies. The fact that stronger individuals do hold such beliefs hints at an evolved psychology specialized for small-scale coalitional aggression, where the strength of coalition members non-negligibly contributes to the net coalition strength. However, most of this research primarily focuses on dyadic aggression, and the results are consistent with alternative explanations that do not require adaptations specialized to war. For example, the fact that stronger individuals endorse the use of military force in state-level conflicts (e.g., Bartusevičius, 2021; Sell *et al.*, 2009; Sell *et al.*, 2017) may reflect the endorsement of force in all types of interactions, irrespective of whether they take dyadic or coalitional forms. In sum, we lack a comprehensive experimental program that probes adaptations *specialized* to war and rules out alternative explanations.

Below, we summarize the hypothesis derivation and testing procedures developed to probe such adaptations. We followed

the hypothesis-generating heuristics commonly utilized in adaptationist research (Lewis *et al.*, 2017: 355–362; see also Williams, 1966; also Mayr, 1983), consisting of the following steps:

1. Theorize the setting of pre-Holocene hominin war.
2. In this setting, identify a major adaptive problem, i.e., an evolutionarily recurrent problem the solution of which promoted reproduction.
3. Theorize an adaptation that contributed to solving this problem.
4. Theorize the specific features of the adaptation via a task analysis.
5. Based on these design features, derive a priori testable hypotheses.
6. Test for the existence of such features, which tests for the existence of the adaptation itself, and addresses alternative explanations.
7. Examine cross-cultural variation in the outputs of the adaptation and address alternative socio-cultural explanations.

Correspondingly:

1. Drawing on archaeological, ethnographic, and primatological literatures referenced above, we inferred that—if hominin war (defined in the generic sense) occurred prior to the Holocene—it most often took the form of hit-and-run ambushes or surprise encounters upon foraging.
2. If such ambushes or encounters regularly occurred in the Pleistocene and earlier, then one of the most consequential problems ancestral hominins faced was how not to fall victim to a coalitional attack upon a surprise intergroup encounter.
3. Solving this problem required mechanisms to rapidly detect and accurately assess potentially threatening outgroup coalitions. We focus on these, and not on other plausible mechanisms (e.g., vigilance at the outgroup foraging range), because they likely constitute specializations to war—a point we elaborate below.
4. An evolutionary task analysis starts by defining the end state and proceeds in specifying the concrete tasks involved in achieving that end state (Marr, 1982). In our case, the end state is surviving a surprise encounter with a potentially threatening group. Achieving this requires, inter alia, distinguishing between animate and inanimate objects, as well as between non-human animals and humans; however, the mechanisms that solve these differentiation problems are not specialized to the domain of coalitional aggression. Below are the basic tasks that we theorized are specific to the above-defined end state of surviving a surprise intergroup encounter:
    i. Detect the outgroup coalition.
    ii. Rapidly and accurately enumerate all members of the outgroup coalition.
    iii. Assess the situational coalitional formidability of the outgroup using variables that evolutionarily determined the outcomes of intergroup fights:
        a. N of outgroup members
        b. Formidability of individual outgroup members
        c. One's own formidability
        d. Availability of ingroup coalitional support

We refer to Tasks I–III, respectively, as the detection, enumeration, and assessment (or computation) tasks. The subsequent behavioral tasks—such as fleeing, standing ground, or attacking—are beyond our research focus. Tasks I–III are not exhaustive, even in the pre-behavior stage, and each task could be subdivided into more specific sub-tasks. We focus here on the minimal set of specialized tasks that we see as necessary, but not sufficient, to achieve the above-defined end state.

Upon certain constellation of factors (e.g., a solitary female foraging with an infant), an immediate flight was likely the most fitness-promoting response in a surprise encounter with potentially threatening outgroups, making the enumeration and computation tasks not only redundant but also potentially costly due to loss of time. Upon other constellation of factors (e.g., male coalitions foraging in resourceful territories), important stakes were likely involved—territories, resources, and status (Wilson & Wrangham, 2003: 363)—requiring more elaborate, yet rapid, weighting of non-flight responses, for which the enumeration and computation tasks would likely be essential.

Completing Task I and II involves visual attention and enumeration components that track specific types of objects in the environment and efficiently enumerate them. Completing Task III includes several computational components that take the inputs from the attention and enumeration components (e.g., N of outgroup members), apply particular algorithms (e.g., "if, then" rules), and—combined—produce an aggregate estimate, referred here to as the situational coalitional formidability index (sCFI). This index can be understood as an internal regulatory variable (Lieberman *et al.*, 2007) which feeds into the subsequent representational and behavioral components. The computations that produce the sCFI are theorized to be automatic and rapid, as they must inform behavior, given the context of a surprise intergroup encounter, unambiguously and swiftly; yet they are also assumed to be complex, assigning different weights and algorithms to multiple inputs, ultimately producing a single summary index. This index is then theorized to be accompanied by a mental representation

(Fessler & Holbrook, 2013b) (e.g., [in the present situation] "outgroup is more formidable than ingroup"). The computations are not directly observable, but we can manipulate the inputs, observe the outputs, and assess whether the inputs produce the outputs in line with the theoretical model. Mental representations are also not directly observable, but we can measure their proxies—people's subjective assessments. Figure 1 presents a visual summary of these theoretical ideas.

We see the three components as neurocognitive mechanisms or circuits that constitute only a part of a larger adaptation, a neurocognitive system evolved to regulate behavior in the specific domain of war. Studying other mechanisms composing the adaptation, especially those pertaining to behavior, would require an empirical strategy different from the one we delineate below. We did not aim to identify all potential mechanisms composing the adaptation but to probe those that would allow inferring the existence of the larger adaptation and—by extension—the existence of pre-Holocene small-scale war. We posit, and elaborate in Discussion, that evidence in favor or against the existence of the detection, enumeration, and assessment mechanisms, as delineated above, would constitute a solid basis to infer the existence or absence of human adaptation to war.

5. The task analysis gave rise to the following hypotheses:

- Task I: People attend, automatically, to potentially threatening coalitions (H1)
- Task II: People enumerate, accurately (H2) and rapidly (H3), members of potentially threatening coalitions
- Task III_a: Formidability assessments of potentially threatening coalitions are a positive function of their numerical size (H4)
- Task III_b: Formidability assessments of potentially threatening coalitions are a positive function of the formidability of their individual members (H5), but this function diminishes with the numerical size of the outgroup coalitions (H6)
- Task III_c: Formidability assessments of potentially threatening coalitions are a negative function of one's own formidability (H7)
- Task III_d: Formidability assessments of potentially threatening coalitions are a negative function of perceived availability of ingroup coalitional support (H8)

H1–H8 are analogous to hypotheses tested in research on formidability assessments in the interpersonal domain. Studies show that people accurately assess the upper-body strength of other individuals in pictures without explicitly being asked to do so and in very short presentation times (as little as, 33ms) (Durkee *et al.*, 2018). Studies also show that one's own strength (Fessler *et al.*, 2014), the presence of allies (Fessler & Holbrook, 2013a), synchronous movement with allies (Fessler & Holbrook, 2014), and social support (Holbrook *et al.*, 2016a) reduce envisioned strength of hypothetical individual adversaries. Pertaining to H5 specifically, one experiment demonstrated that, following a surprise presentation, a schematic group consisting of three muscular individuals and one weaker individual was perceived as stronger compared to a

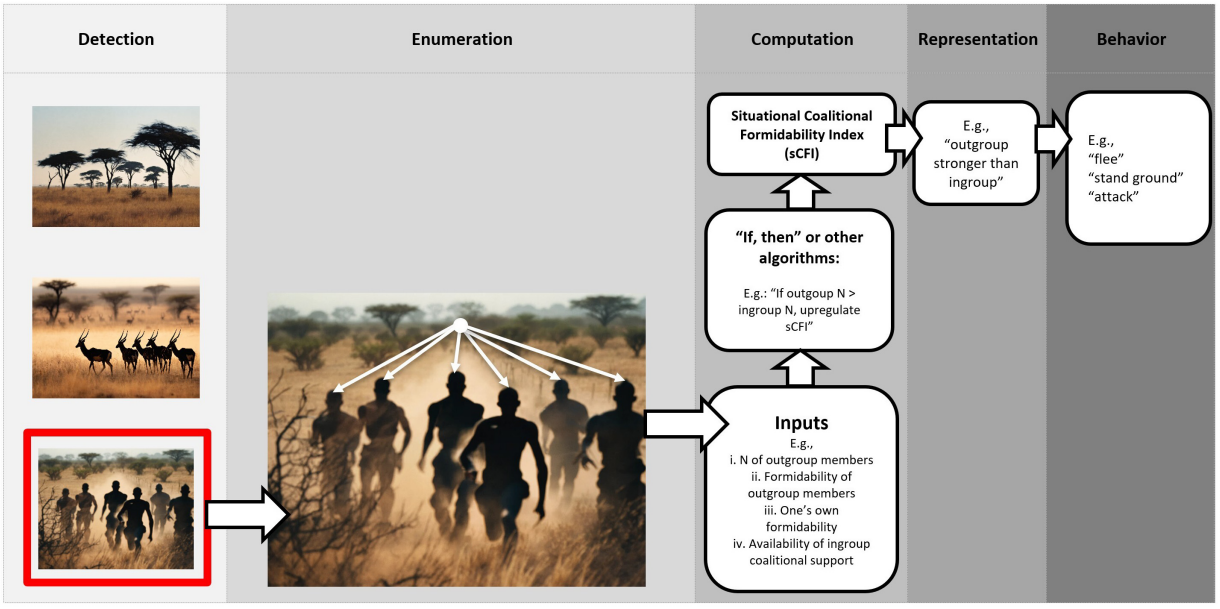



**Figure 1. A theoretical model of an evolved adaptation specialized to regulate behavior in the domain of small-scale war.** Images were generated with Lexica Aperture V4.

group consisting of three weaker individuals and one muscular individual (Durkee *et al.*, 2018).

In addition to the core hypotheses, we specified hypotheses pertaining to sex. Ethnographic and comparative literatures reviewed above suggest that intergroup coalitional aggression is primarily a male activity; if this also characterized pre-Holocene hominin war, then we should expect that:

- Males, compared to females, determine the size of potentially threatening coalitions more accurately (H9) and rapidly (H10)
- When females make errors in these estimates, they tend to overestimate more so than males (H11)
- Females, compared to males, subjectively assess potentially threatening coalitions as more formidable (H12)

H12 draws on the assumption that ancestral females, compared to males, had lower self-defense capacity. Since formidability of outgroups is relative to that of ingroups, lower self-defense capacity among individual ingroup members implies higher formidability of outgroups. H11 draws on the assumption that in the context of a surprise encounter with a potentially threatening outgroup, upholding resources and status were of less value for females than males (Wilson & Wrangham, 2003: 363). Females were thus more likely to flee, potentially skipping the enumeration and computation steps, defaulting to the assumption of greater outgroup formidability. Following this logic, individuals with lower self-defense capacity, given the limited time to gather formidability information, would also tend to err on the side of caution, assuming more formidability in the outgroup. If this holds true, then self-defense capacity as such should influence coalitional assessments. Self-defense capacity is a function of multiple variables. The key among these are individual formidability and availability of ingroup coalitional support. Hence, we also hypothesize that:

- Less formidable individuals tend to overestimate the size of potentially threatening coalitions (i.e., make positive errors in the estimates) (H13)
- Individuals with limited access to ingroup coalitional support tend to overestimate the size of potentially threatening coalitions (i.e., make positive errors in the estimates) (H14)

"Potentially threatening coalitions" are defined here as those characterized by features that evolutionarily posed the threat of a coalitional attack. The core of these are the male sex and coalitional coordination. If war recurred during human evolution, and if it typically took the form of coordinated groups of males attacking other groups in a surprise fashion, then our evolved minds likely associate coordinated groups of unfamiliar males with the threat of a coalitional attack. Correspondingly, our stimuli rating studies (presented below) revealed that both, features typical of the male sex and postural coordination, correlated with coalitionality perceptions—the extent to which participants perceived schematic representations of humans as forming coalitions. These perceptions, in turn, correlated with threat perceptions, implying that groups of individuals who are perceived as coalitions are also perceived as threatening. Importantly, the correlation between coalitionality and threat perceptions was stronger for male-like (vs female-like) targets.

Coalition detection, enumeration, and assessment mechanisms are likely only activated in the context of intergroup conflict. The appearance of ingroup males in the visual field should not activate mechanisms that are specialized to the domain of intergroup coalitional aggression. Even outgroup males may not activate these mechanisms if ingroup-outgroup relations had been cooperative. This implies a potential moderating effect of the intergroup context. In the main experiments, we do not explicitly model such moderators (but see Discussion). We assume that the above-delineated mechanisms operate on the fire alarm principle: in the absence of conflict, or when outgroup's status is unknown (hostile vs neutral vs cooperative), it is less costly to assume a threat of a coalitional attack and to act accordingly, compared to assuming no threat and risking injury or death. Our approach is thus conservative: if we find that the theorized mechanisms are activated in contextless situations, then we can infer—with greater certainty—the existence of the adaptation, which should be activated more strongly in explicit contexts of intergroup conflict.

Multiple other variables could be theorized to influence coalitional assessments, such as presence of weapons or familiarity with the terrain at the encounter (Kelly, 2005). However, our goal was not to produce an exhaustive list; rather, we aimed to specify hypotheses that allow designing an experimental program that would constitute a strong test of the overarching hypothesis that humans are equipped with specialized adaptation to small-scale war. Furthermore, although we acknowledge the importance of weapons and situational variables, we assume that—ceteris paribus—coalitional formidability was a strong function of the numerical size and formidability of individual members. Accordingly, we assume that adaptations specialized to intergroup coalitional aggression used these variables as key inputs in formidability assessments.

We formulated H1–H14 as general hypotheses, which can be tested in various ways. In Design subsection, for each of these, we derive specific predictions matched to the experimental paradigms we used in our own experimental program.

6. We aimed not only to empirically test H1–H14 but also address alternative explanations. If people automatically attend to coordinated groups of males and then rapidly and accurately determine their size and formidability—and if such assessments are moderated by sex, self-defense capacity, and other ancestrally relevant variables—then what other phenomena might account for all this other than adaptations to solve the problem of surviving a surprise encounter with a potentially threatening group? To substantiate an evolved adaptation, one must demonstrate that it is not more parsimoniously explained as a by-product of another adaptation specialized to solve some other problem (Williams, 1966). Automatic attention to coordinated males may be a by-product of a more general tendency

to attend to humans (as contrasted to other objects); or, for instance, a by-product of a tendency to attend to aligned objects, whether human or not. Likewise, people may efficiently enumerate coordinated male-like groups due to the similarity grouping effect (Yu *et al.*, 2019), not because coordinated males as such cue for threat that requires rapid appraisal. Our experimental program has been designed and piloted to address these and other alternatives.

7. Evolved psychological mechanisms are assumed to be universal, reliably developing in all typical individuals (Cosmides & Tooby, 2015); yet they are not insensitive to environmental inputs. An adaptive psychological mechanism must produce specialized behavioral outputs that are adaptive in specific types of recurrent situations. The first goal of Step 7 is to test whether coalition detection, enumeration, and assessments vary across different countries or cultures. The second goal is to explain such variation. For example, people may efficiently enumerate outgroup coalitions in countries that historically have had a higher incidence of violent intergroup conflict. To pursue these goals, we examine whether the effects hypothesized above significantly vary across 30 countries and in turn explore country and sub-country variables that account for this variation.

Importantly, in Step 7 we also explicitly test alternative socio-cultural explanations for the results from Step 6. For example, the finding that males, compared to females, produce more accurate assessments of coalitions (H9), may be a result of prevalent gender roles, not because males evolutionarily engaged more often in warfare. Given the breadth of the project, Step 7 will be reported in a separate publication. Although individual steps are reported in separate publications, the project follows an integrated, multi-step design (a point we return to in Discussion).

## Materials and methods

### Pilot data

A total of 25 exploratory studies, hereby "pilots" (P1, P2.1–2.8, P3.1–3.16), with 5,222 participants were conducted (SI1, SI2, and SI3, respectively, in the Supporting Information[1]), aimed at (a) developing experimental stimuli, (b) developing experimental paradigms, and (c) preliminarily testing H1–H14.

#### *Developing experimental stimuli*

To match the simulated setting of a surprise intergroup encounter, we used images for stimuli. We started with schematic black silhouettes, modelled to match those in existing research on related topics (pp. 11–12 in SI1.1), and then shifted to 3D images to enhance realism (pp. 11–12 in SI2.1). We used default body shapes in the Magic Poser Webapp and then manipulated features to enhance characteristics expected to produce the hypothesized effects: postural coordination and morphological features typical of the male vs female sex, such as shoulder breadth, hip-to-waist ratio, and upper-body musculature (Figure 2 and Figure 3). Given the short presentation times and schematic targets, we pronounced these features rather than modelling them to approximate some real-world references. We did not aim to identify shapes that represent some average and culturally appropriate individuals but to produce stimuli in which coordination vs non-coordination and male vs female contrasts were easily and rapidly discernable across cultures—at the same time minimizing extraneous features that might evoke unknown responses in participants.

We opted for schematic, rather than naturalistic, stimuli for the following reasons. Naturalistic images, e.g., photos of Americans, contain culture-specific characteristics, such as clothing or hairstyle, which might influence attention and other processes in participants in unknown ways. Attentional bias toward a coordinated group of men, dressed in some specific way, may be due to clothing details, and not due to threat, thereby confounding results. Such details could in principle be held constant across the experimental and control stimuli, but with naturalistic images such control is difficult, especially for male-female contrasts. Furthermore, presenting pictures of people typical to some country to participants from other countries may differentially engage their attention or evoke other unknown responses, raising measurement equivalence concerns.

We asked Prolific participants to rate the stimuli with different target dispersion, N, and size, on two dimensions: coalitionality—the degree to which one perceives collectives of individuals as forming coalitions ("a group of individuals who act together to achieve a goal"); and threat—the degree to which one finds the images threatening. A snapshot of these ratings, with mean scores represented by red lines, is presented in Figure 2 and Figure 3 (P2.1–2.6 and P2.7–2.8 used different sets of rating pictures). Coalitionality ratings revealed that:

- simple horizontal rotation of some targets produced large increases in coalitionality ratings;
- coordinated male-like targets, compared to coordinated female-like targets, received slightly, but significantly, higher coalitionality ratings;
- increasing the number of coordinated male-like targets increased coalitionality ratings;
- changing one-rowed formation to two-rowed further increased coalitionality ratings.

Considering threat perceptions:

- simple horizontal rotation of some targets produced notable changes in threat perceptions;
- increasing the N of coordinated male-like targets increased threat ratings;
- changing one-rowed formation to two-rowed further increased threat ratings;
- coalitionality and threat ratings correlated, more so for male-like targets than for female-like targets.

We also examined how participants rated the same targets but distributed in rectangular arrays (P3.1–P3.16 presented stimuli

[1] Supporting Information is available on the Open Science Framework (OSF) repository associated with the project: https://doi.org/10.17605/OSF.IO/RYV4N

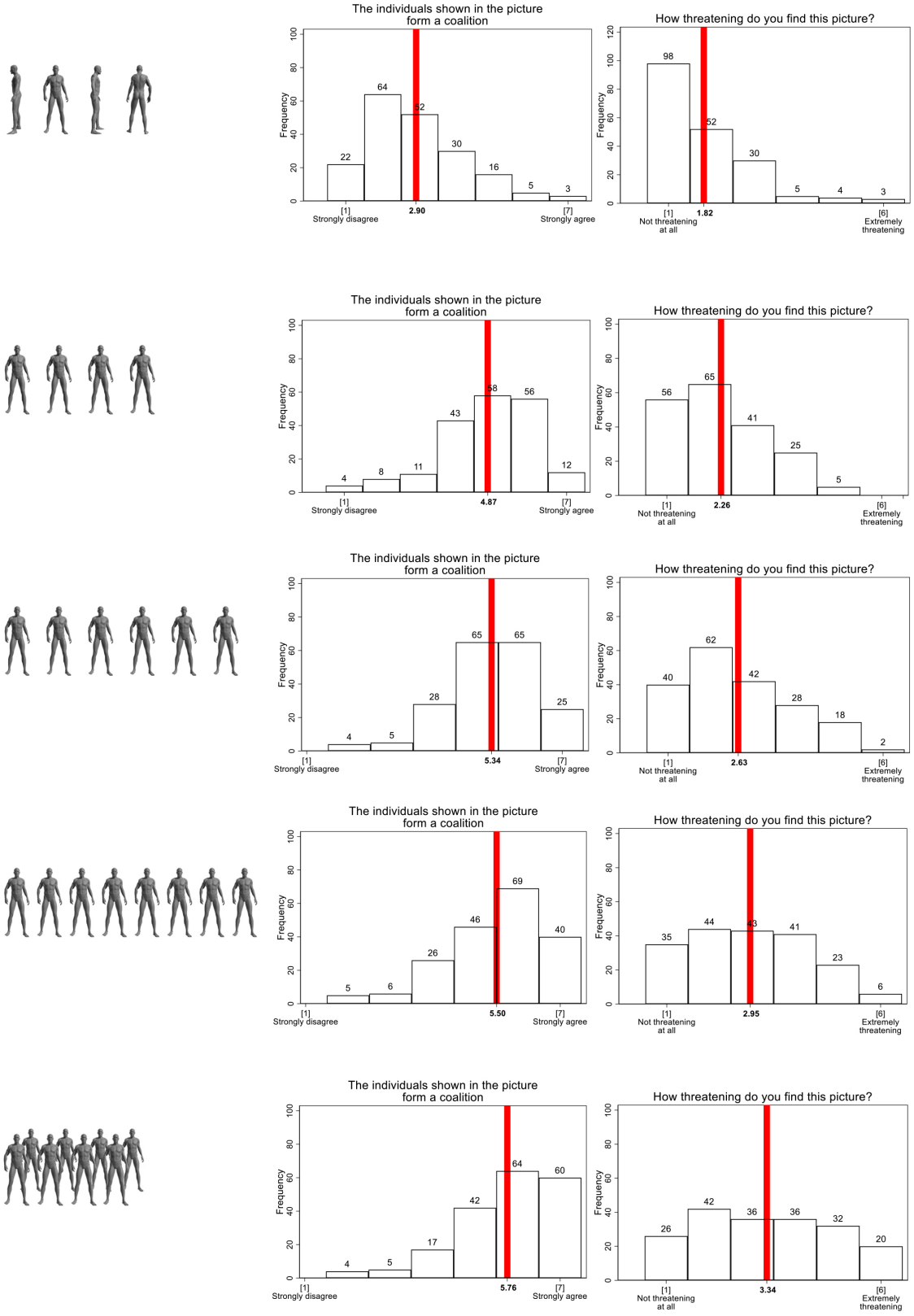



**Figure 2. Coalitionality and threat ratings in Pilots 2.7 and 2.8.** Red vertical lines represent mean scores.

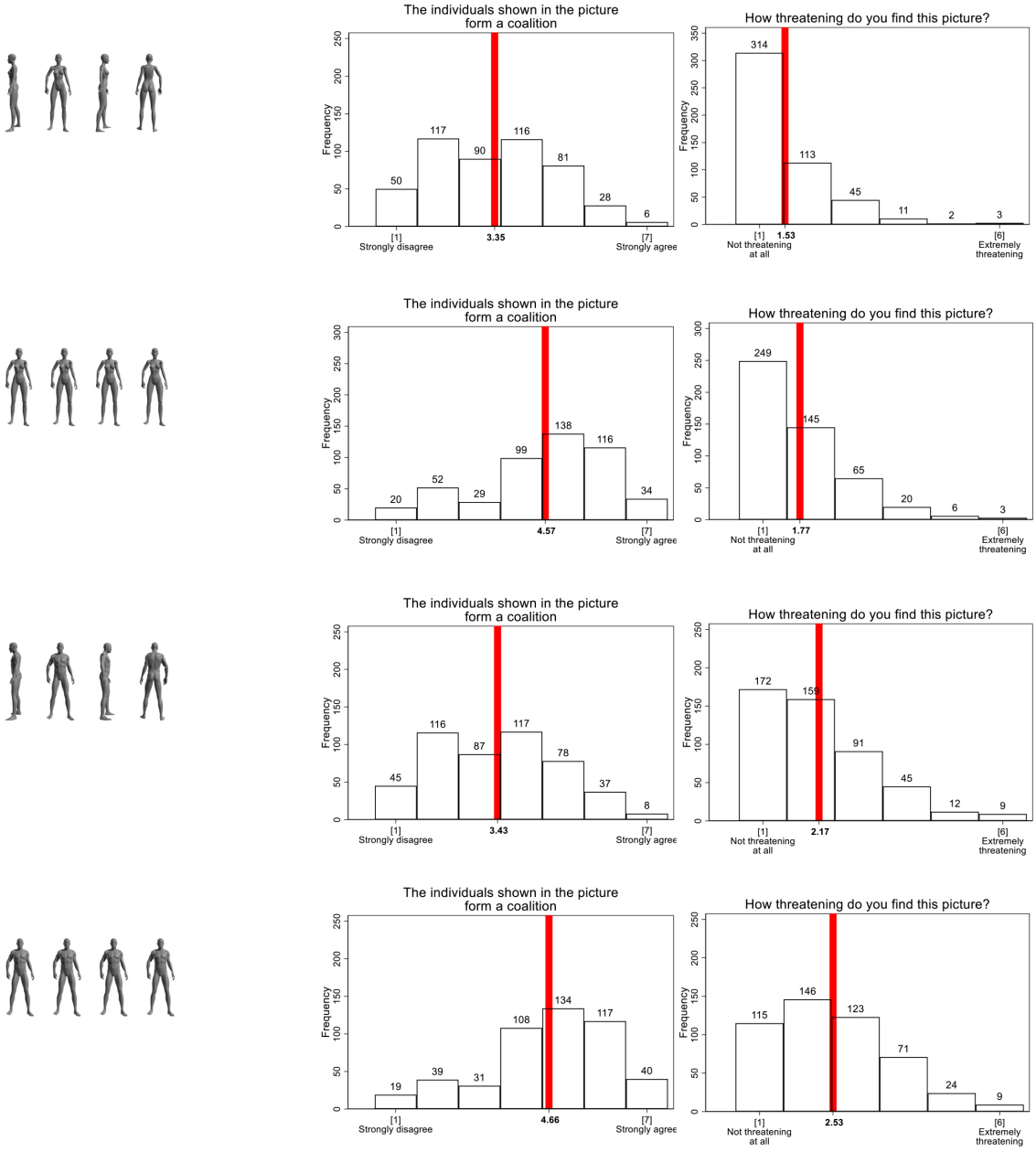



**Figure 3. Coalitionality and threat ratings in Pilots 2.1–2.6.** Red vertical lines represent mean scores.

in rectangular arrays). P3.1–P3.10 revealed that increasing target size in arrays produced higher threat ratings (p. 30 in SI3.2).

P3.11 and P3.12 used subjective ratings of stimuli as the main task, for which we used images from P2.8 as the main stimuli.

***Developing experimental paradigms***

We used surprise-recall (P1; 2.1–2.8), dot-probe (P1; 2.1–2.8), object-enumeration (P3.1–3.16), and rating tasks (P3.11 and P3.12) to test our hypotheses. The surprise-recall task was discontinued (see pp. 42–43 in SI2.2).

The dot probe paradigm (Macleod *et al.*, 1986) is a widely used task in cognitive psychology to study automatic attention processes (Thigpen *et al.*, 2018), which we used to assess whether people automatically attend to potentially threatening coalitions. Attentional processes are subdivided into several categories (Cisler & Koster, 2010). Given our goal to simulate a surprise intergroup encounter, we focused on the initial shift

in attention toward the experimental stimuli, i.e., attentional orienting. Attentional orienting can be exogenous (automatic) or endogenous (intentional) (McCormick, 1997), and our substantive focus was on the former. Given this, we used short presentation times. Studies show that measuring attentional orienting, specifically, improves reliability of the dot-probe task (Chapman *et al.*, 2019). Extant literatures do not provide exact recommendations for stimuli presentation times aimed at capturing exogenous attentional orienting. At least one study demonstrated the highest internal reliability at 100ms exposures (Chapman *et al.*, 2019). Across P2.1–2.8 we also found 100ms to reveal, consistently, the hypothesized orienting toward the experimental stimuli (pp. 39–41 in SI2.2).

In the standard form, the dot-probe trials start with a presentation of a fixation cross on the center of a screen, followed by a pair of visual stimuli, one presented on the left, and another on the right, side of the screen. Subsequently, the stimuli pair disappears and is followed by a symbol probe X, either on the left or the right side of the screen. The participants are instructed to respond to the X by clicking—as quickly as possible—one of two keys on the keyboard: N if the X appears on the left and M if the X appears on the right. We introduced a modification to this standard form, suggested in extant research (Staugard, 2009), and examined in Pilots 2.7 and 2.8 (p. 41 in SI2.2). Upon appearance of X, we programmed the task to retain the pictures on the screen. After several trials, learning that the appearance of X coincides with the disappearance of pictures, the participants may only start engaging with the task after the disappearance of the pictures. An eye-tracking study confirmed this pattern, finding that participants tend to only initiate their search when the probe appears (Mogg *et al.*, 2000).

We also examined pairs that contained neutral objects and scrambled inverted versions of the target stimuli as control stimuli. These contrasts were discontinued (see p. 13, SI1.2). The control in our dot-probe experiments is thus served by the contrast image in each pair with identical secondary features (e.g., color and size) but representing different sex, postural rotation, or N of targets.

The dot enumeration paradigm is a widely used task to study numerical cognition (Schleifer & Landerl, 2011), which we used to assess whether people enumerate potentially threatening coalitions more efficiently, and whether such enumeration relates to raters characteristics, such as sex and formidability. The standard design displays different Ns of dots and asks the participants to count them as fast as possible. We piloted two modifications of this task: estimation accuracy and counting rapidity, assessing (respectively) how accurately and rapidly participants estimate target N. Following the standard design, instead of rows, we used square arrays with randomly distributed targets. We tested different sizes of the arrays and targets, finding that the hypothesized effects were weaker for smaller targets, presumably because of the difficulty to distinguish sex differences and postural coordination. We therefore scaled up the target images, simultaneously reducing the array size, eventually settling on an array size of 3*5.

We then examined different target numbers. Drawing on research reviewed in Introduction, suggesting that raiding or foraging parties were of relatively small size, we delimited our pilots to 3–10 targets. We found that target category effects were consistently dimmed or non-existent in the subitizing range (3–4), presumably because such small Ns were easy to enumerate, producing ceiling effects. Conversely, we found that the task difficulty, as indicated by N of targets, appeared to reveal stronger hypothesized effects (pp. 30–31 in SI3.2). A technical explanation of this pertains to ceiling effects. An alternative, substantive interpretation is that mechanisms evolved to navigate coalitional aggression operate more efficiently, compared to other mechanisms (e.g., general counting ability), under extreme conditions. Although we found that the hypothesized effects became stronger with increasing target N, this seemed to cap out at 8 or 9. Presumably, this was due to the 3*5 arrays becoming too crowded with 9 or more targets, aiding enumeration irrespective of target category. Taken together, this led us to settle on the target range of 6–8.

For the estimation accuracy version, we tested several stimuli presentation times (200ms, 400ms, 500ms), eventually settling on 400ms, as that produced moderate accuracy rates not conducive to ceiling effects (pp. 28–29 in SI3.2). Finally, we examined how cues of ingroup/outgroup membership assigned to targets influence enumeration performance (P3.13–P3.16), but this manipulation was discontinued for the reasons explained on p. 37 in SI3.2.

The rating task, piloted in P3.11 and 3.12, was developed based on the ratings provided throughout P2.1–2.8. In Pilots 3.11 and 3.12, we exposed participants to a series of stimuli with coordinated male-like targets identical to those in P2.7 and P2.8 (Figure 2) but expanded the target N to a 4–10 range. Participants were asked to assess the fighting capacity of the would-be coalitions composed of the targets. Drawing on existing research suggesting that body size is a key evolutionary cue of individual formidability (Fessler & Holbrook, 2014; Fessler & Holbrook, 2016; Fessler *et al.*, 2016), we also manipulated the size of one of the targets to assess the effect of individual formidability on the net coalition fighting capacity.

In sum, the final designs of the experimental tasks are based on Pilots 2.7-8 for the dot-probe paradigm, P3.6 for the counting rapidity, P3.10 for the estimation accuracy, and P3.11–12 for the coalitional formidability rating task.

***Preliminary tests of H1–H14***

Below is our assessment of empirical support for each hypothesis, combining results from multiple tests across the pilots. These findings reflect overall patterns rather than results from individual tests (for test-specific estimates, please refer to the indicated pages):

- *Hypothesis 1 was supported with low uncertainty* (Pilots 2.7, 2.8, pp. 39–41, SI2.2): Participants attended to coordinated male-like targets when contrasted with

identical non-coordinated male-like targets, but not to coordinated female-like targets when contrasted with non-coordinated female-like counterparts. This suggests that attention is drawn not to coordination per se, but specifically to coordination among male-like targets—presumably because coordinated groups of males cue for threat of a coalitional attack.

- *Hypothesis 2 was supported with low uncertainty* (Pilot 3.10, pp. 28–29, SI3.2): Participants enumerated coordinated male-like targets more accurately than rotated male-like or coordinated female-like targets.
- *Hypothesis 3 was supported with moderate uncertainty* (Pilot 3.6, pp. 20–21, SI3.2): Participants enumerated coordinated male-like targets more rapidly than rotated male-like or coordinated female-like targets.
- *Hypothesis 4 was supported with low uncertainty* (Pilot 3.11, 3.12, pp. 35, SI3.2): Participants rated larger groups of male-like targets as higher in coalitional formidability.
- *Hypothesis 5 was supported with low uncertainty* (Pilot 3.11, 3.12, pp. 35, SI3.2): Participants rated groups of male-like targets that included a single larger individual as having greater coalitional formidability than otherwise identical groups composed of uniformly sized individuals.
- *Hypothesis 6 was supported with low uncertainty* (Pilot 3.11, 3.12, pp. 35, SI3.2): The impact of a standout individual on coalitional formidability ratings diminished as group size increased, evidenced by a negative interaction between the presence of a single large individual and the number of targets in the group.
- *Hypothesis 7 was supported with moderate uncertainty* (Pilot 3.12, pp. 38, SI3.2): Individuals with lower self-assessed formidability rated coordinated male-like targets as having greater coalitional formidability.
- *Hypothesis 8 was not supported* (Pilot 3.10, pp. 38, SI3.2): Contrary to expectations, individuals with lower perceived ingroup coalitional support did not assign higher coalitional formidability to the same targets.
- *Hypothesis 9 was supported with moderate uncertainty* (Pilot 3.10, pp. 38, SI3.2): Male participants enumerated coordinated male-like targets more accurately than female participants.
- *Hypothesis 10 was supported with low uncertainty* (Pilot 3.6, pp. 38, SI3.2): Male participants enumerated coordinated male-like targets more rapidly than female participants.
- *Hypothesis 11 was supported with moderate uncertainty* (Pilot 3.10, pp. 38, SI3.2): Female participants tended to overestimate the number of coordinated male-like targets.
- *Hypothesis 12 was supported with low uncertainty* (Pilots 3.11, 3.12, pp. 38, SI3.2): Female participants also rated coordinated male-like targets as more formidable than did male participants.
- *Hypothesis 13 was not supported* (Pilot 3.10, pp. 38, SI3.2): By contrast, individuals lower in formidability did not exhibit similar overestimation errors, contrary to expectations.
- *Hypothesis 14 was not supported* (Pilot 3.10, pp. 38, SI3.2): Those with lower perceived coalitional support also did not exhibit similar overestimation errors, contrary to predictions.

Rather than rejecting the hypotheses that received no support at the exploratory state, and instead of corroborating those that did, we set out to more comprehensively test them in the main confirmatory studies. The pilots involved arbitrary modelling choices, which were not preregistered. Furthermore, the samples in some pilots were small, implying an increased risk of type 1 errors.

## Finalized preregistered designs

We followed a conservative hypothesis testing strategy. The theorized adaptations that evolved to navigate intergroup coalitional aggression, would be activated most potently by naturalistic cues in the context of hostile intergroup conflict. If such adaptations are activated by highly schematic cues, and in participants who complete experimental tasks on computers in safe, modern environments, then they would highly likely be activated in real world contexts in individuals encountering members of hostile outgroups.

The finalized experimental program was composed of four types of designs:

1. Automatic attention experiments (Studies 1 and 2)
2. Estimation accuracy experiments (Studies 3 and 4)
3. Counting rapidity experiments (Studies 5 and 6)
4. Subjective formidability experiments (Study 7)
5. Cross-cultural experiments (Study 8; presented in a separate publication)

In what follows, we overview each design type, building on the pilots reported above. Complete protocols of Studies 1–7, based on SEPTRE standards, are available in SI (SI4.1 and SI4.2).

***Studies 1 and 2: Automatic attention experiments***

The automatic attention experiments utilized the standard dot-probe paradigm with the modifications as described in Pilot Data. For the stimuli, we took the final versions as rated

in P2.7 and P2.8, corrected minor morphological details, and enhanced their quality with SDLX Turbo by Stability AI (Figure 4). We used five categories (for higher quality images see pp. 11–12 in SI4.1):

1. Four coordinated targets with pronounced features typical of the male sex.
2. Four rotated targets with pronounced features typical of the male sex.
3. Four coordinated targets with pronounced features typical of the female sex.
4. Four rotated targets with pronounced features typical of the female sex.
5. Eight coordinated targets with pronounced features typical of the male sex in two-row formation.

The four stimuli were contrasted with each other. A task that has many trials increases the chance of participant disengagement and desensitization to the stimuli. Our pilots have shown that trial N diminishes the hypothesized category effects (pp. 31–34 in SI3.2). Some participants reported that they found the tasks with large N of trials monotonous (p. 7 in SI1.2). Theoretically, we also see limited exposures as more matching our simulated setting: an adaptation that evolved to navigate surprise intergroup encounters may not be activated, or not to the same extent, when the same stimuli are repeated many times. Therefore, we split the dot-probe trials into two sets, administered separately in Studies 1 and 2.

In Study 1, we presented the following combinations (conditions):

- Combination 1: coordinated male-like group vs rotated male-like group;
- Combination 2: coordinated male-like group vs large coordinated male-like group;
- Combination 3: rotated male-like group vs large coordinated male-like group.

In Study 2:

- Combination 4: coordinated female-like group vs rotated female-like group
- Combination 5: coordinated female-like group vs coordinated male-like group
- Combination 6: coordinated female-like group vs large coordinated male-like groups

To counterbalance, following the standard design, each pair was presented four times, with the presentation order randomized:

1. Both the target stimulus (e.g., coordinated male-like group) and X on the left side of the screen (left side congruent trial).
2. The target stimulus on the left side, but X on the right side, of the screen (left side incongruent trial).
3. Both the target stimulus and X on the right side of the screen (right side congruent trial).
4. The target stimulus on the right side, but X on the left side, of the screen (right side incongruent trial).

Finally, to improve measurement, each presentation was repeated twice. The dot-probe task thus was composed of 3 x 4 x 2 = 24 trials in each Study 1 and 2. The participants started by completing several practice trials, which require satisfactory performance to proceed to the main trials (p. 23 in SI4.1). Combined with informed consent (pp. 35–36 in SI4.1),

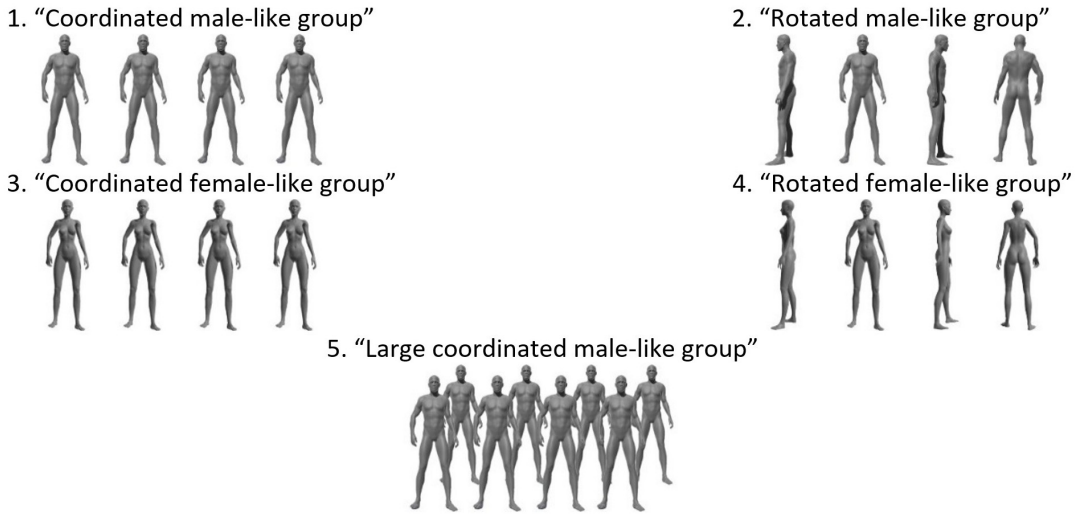



**Figure 4. Stimuli used in Studies 1 and 2.**

instructions (p. 23 in SI4.1), training trials, and post-task questionnaire (pp. 29–30 in SI4.1), the task took on average approximately 8 minutes to complete.

Reaction time (RT)—milliseconds elapsed from the onset of the probe X until the M- or N-click were recorded. Shorter RTs imply that participants attend to visual stimuli that appear on the same side of the screen as the X, and longer RTs imply that participants attend to visual stimuli that appear on the other side. The participants were not instructed to focus on any of the visual stimuli, only on the fixation cross, and then click as fast as possible keys once X appeared. Hence the task allowed assessing what participants attend to automatically.

Correspondence to hypotheses. Designed in this way, the dot-probe task allows testing the hypothesis that people automatically attend to potentially threatening coalitions (H1). As described above, "potentially threatening coalitions" are understood here as coordinated groups of males. In what follows, we specify concrete predictions derived from H1, as well as potential alternative explanations, matched to this specific paradigm.

Study 1 tested whether participants automatically attend to coordinated male-like groups when contrasted to rotated male-like groups that are otherwise identical (Combination 1) (Prediction 1); and whether participants automatically attend to larger coordinated male-like groups when contrasted to smaller coordinated (Combination 2) (Prediction 2) and rotated (Combination 3) (Prediction 3) male-like groups that are otherwise identical.

Automatic attention to coordinated male-like groups may be due to a more general psychology that attends to all humans. The contrast images in each trial allow addressing such alternatives, because they all contain human-like targets. Participants may also attend to coordinated male-like groups because they contain male-like targets with pronounced features typical of the male sex, such as upper body musculature, which may themselves cue for threat. The contrast with rotated male-like targets allows addressing this alternative: if people attend to solitary males in coalitions, and not to coalitions as such, then our tests would reveal no attentional bias toward coordinated male-like targets, when contrasted to identical but rotated targets.

Participants may attend to coordinated male-like groups, when contrasted to identical but rotated groups, because postural coordination as such attracts attention, not because coordination serves as a cue of threat of a coalitional attack. If postural coordination as such attracts attention, then participants will attend to coordinated female-like groups when contrasted to rotated female-like groups (Combination 4). If not, or if the attentional bias toward coordinated female-like groups is weaker, then coordination only attracts attention when the coordinated figures are male-like, presumably because coordinated males, but not coordinated females, evolutionarily constituted a threat of a coalitional attack.

Participants may attend to male-like groups when contrasted to female-like groups (Combination 5), but this attentional bias is likely diminished because the attractiveness of stimuli may also draw attention. Our pilot data revealed that female participants, compared to male, attended more to male-like stimuli; whereas male participants, compared to female, attend more to female-like stimuli. However, stronger cues of threat, as represented by larger coordinated male-like groups, potentially overpower cues of attraction (Combination 6).

We only preregistered formal tests for Combinations 1–3 (Predictions 1–3), considering tests of Combinations 4–6 as exploratory, concerned with alternative explanations. Combinations 1–6 each constitute separate sub-experiments, each with one target stimulus and one control stimulus. For example, in the coordinated male-like vs rotated male-like pair, the latter serves as a control for the former, holding "humanness", "maleness", N, and secondary visual features (such as target size and color)—but not coordination—constant. Accordingly, the data from each condition was analyzed separately. The stimuli pairs were randomized both within and between participants, but the analyses focused on the within-individual variation; see Analysis Plan.

***Studies 3 and 4: Estimation accuracy experiments***

The estimation accuracy experiments utilized the adapted version of the dot enumeration paradigm as described in Pilot Data. The trials began with a fixation cross presented at the center of the screen for 1,500ms. Subsequently, an enumeration array appeared and disappeared after 400ms, followed by selectable response options ranging from 1 to 9, asking participants to estimate—as accurately as possible—the N of targets in the previously shown array. After the numerosity input, the next trial began, starting with the fixation cross and repeating the procedure as described above. Participants completed several practice trials prior to the main task, requiring satisfactory performance to proceed to the main trials.

The enumeration arrays contained 6 to 8 targets. We used identical targets to those used in Studies 1 and 2 but dispersed across 3*5 arrays (Figure 5). The dispersion of targets across the 3*5 arrays was randomized. In addition, we included another target category—black dots—to capture baseline enumeration accuracy across participants. The estimation accuracy pilots have shown that trial N considerably diminishes the hypothesized category effects (pp. 31–34 in SI3.2). Therefore, we also split the object-enumeration trials into two sets, administered separately in Studies 3 and 4.

In Study 3, we present the following stimuli:

- Coordinated male-like groups in arrays

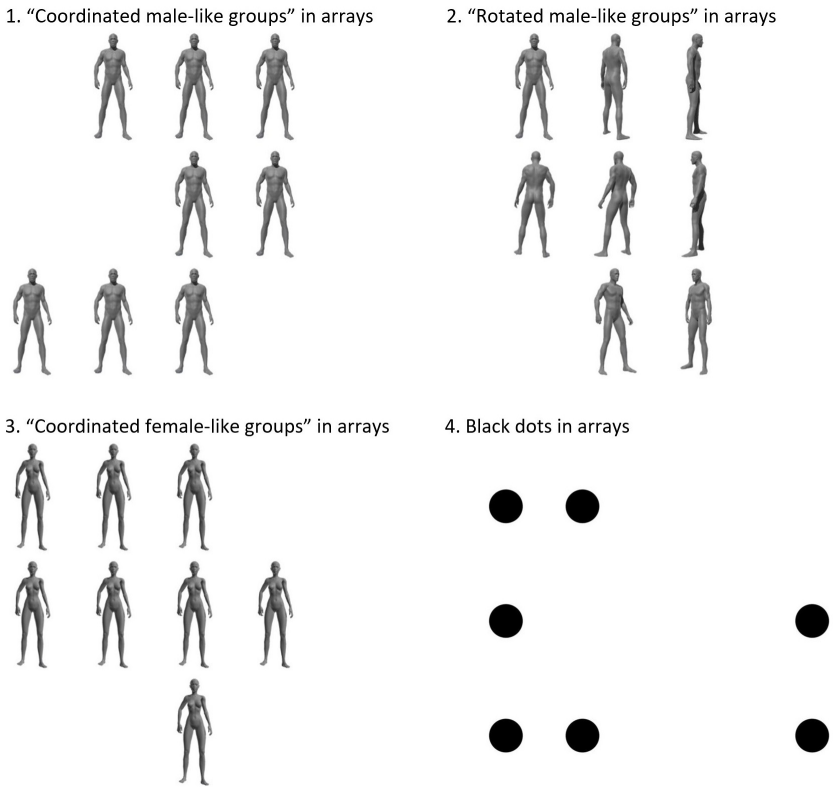



**Figure 5. Stimuli used in Studies 3–6.**

- Rotated male-like groups in arrays
- Black dots in arrays

In Study 4, we presented the following stimuli:

- Coordinated male-like groups in arrays
- Coordinated female-like groups in arrays
- Black dots in arrays

To improve measurement, each stimulus, for each numerosity, was repeated three times. The task thus consisted of 3 x 3 x 3 = 27 trials in each Study 3 and 4. The order of the 27 trials was randomized within and between individuals, but the analyses focused on the within-individual variation; see Analysis Plan. Estimation accuracy was represented as the difference between the selected response and the actual target N displayed in each trial. The outcome variable thus takes the value 0 if the participant was correct, a negative value if the participant underestimated the target N, and a positive value if the participant overestimated the target N. As an alternative, we also used a binary indicator of correct response.

Correspondence to hypotheses. Studies 3 and 4 allowed testing whether participants accurately estimate the size of potentially threatening coalitions (H2). Matched to this specific paradigm, H2 gave rise to the following predictions: coordinated male-like targets, compared to rotated male-like targets (Prediction 4) and coordinated female-like targets (Prediction 5), are estimated with lower error. Due to the similarity grouping effect, alike objects may be easier to enumerate and so support for Prediction 4 may not necessarily imply that coordinated male-like targets are enumerated better because they cue threat. Therefore, we specified Prediction 5, in which both stimuli category contain aligned targets: if coordinated male-like targets are enumerated better than coordinated female-like targets, then this category effect cannot be due to target similarity. The dot arrays were included for comparability to extant work, which typically use dots as enumeration targets. The pilot data revealed that coordinated male-like targets, but not other human-like targets, were estimated equally accurately as dots, despite the latter’s visual simplicity (see p. 30 in SI3.2). Studies 3 and 4 also allowed testing whether male participants, compared to female, estimate the size of potentially threatening coalitions more accurately (H9); and whether females (H11), less formidable individuals (H13), and those with perceived lower availability of coalitional support (H14) tend to produce overestimation errors. That is, we predicted that male participants, compared to female, would produce smaller estimation errors for coordinated male-like targets (Prediction 6); and that the female sex (Prediction 7), lower self-assessed physical formidability (Prediction 8), and lower perceived

availability of ingroup coalitional support (Prediction 9) would predict overestimation errors for coordinated male-like targets.

***Studies 5 and 6: Counting rapidity experiments***

Studies 5 and 6 were identical to Studies 3 and 4, respectively, aside from the outcome measure. The participants were exposed to the same stimuli set but, instead of disappearing after 400 ms, the stimuli remained on the screen until counted. The participants were instructed to count the targets as fast as possible and press the spacebar, followed by an enumeration array ranging from 1 to 9. Counting rapidity was represented by the latency (ms) from the onset of the stimuli to the spacebar click. We only used correct counts in the analyses. The participants also completed practice trials requiring satisfactory performance.

Correspondence to hypotheses. Studies 5 and 6 allowed testing whether participants more rapidly count the N of individuals composing potentially threatening coalitions (H3). In terms of specific predictions, we expected that coordinated male-like targets, compared to rotated male-like targets (Prediction 10) and coordinated female-like targets (Prediction 11), would be counted faster. Studies 5 and 6 also allowed testing whether males enumerate potentially threatening coalitions faster (H10). That is, we predicted counting latencies to be shorter for male participants, compared to female, when enumerating coordinated male-like targets (Prediction 12).

***Study 7: Subjective formidability experiments (Hypotheses 4–8, 12)***

The participants were asked to subjectively rate a series of images. The images were identical to coordinated male-like groups as in Studies 1–2, but the target N ranged from 4 to 10. Target N served as a continuous predictor. Furthermore, each participant rated 4 to 10 targets of identical size, and 4 to 10 targets of identical size aside from one target that was 15% larger than the others (randomly placed within each row). Each participant thus completed 7 × 2 = 14 trials, with the order of the images randomized. The targets were presented for 400ms, followed by the following question: "On a scale from 1 to 10, please rate the formidability (fighting capacity) of the coalition displayed in the last picture". Prior to this question, a definition of a coalition was given: "A coalition is a group of individuals who act together to achieve a goal". Figure 6 presents examples of stimuli, using a group size of 10 targets, with and without one large target.

Correspondence to hypotheses. Study 7 allowed assessing how N of individuals in a coalition influences subjective formidability ratings of the coalition (H4), how formidability of individual members influences the net formidability ratings of coalitions (H5), and whether the latter effect diminishes with the increasing size of the coalitions (H6). Based on prior research, we cued for formidability with body size—a key, and phylogenetically ancient, cue of individual formidability (Fessler & Holbrook, 2014; Fessler & Holbrook, 2016; Fessler *et al.*, 2016). For concrete predictions, we expected that coordinated male-like targets with larger target N (Prediction 13) and those with one larger individual (Prediction 14), would be rated as having higher coalitional fighting capacity. However, we expected the effect of large individual to diminish with numerical size of such coalitions, i.e., a negative multiplicative interaction between a dummy indicating a presence of a large individual and a continuous variable representing the numerical size of the coalition (Prediction 15). Study 7 also allowed assessing whether self-perceived formidability of the raters (H7), perceptions about ingroup coalitional support (H8), and raters' sex (H12) correlate with outgroup formidability ratings. Self-perceived formidability and availability of ingroup coalitional support was measured using the post-task questionnaire, described below. We thus predicted that participants with self-assessed lower formidability (Prediction 16), those

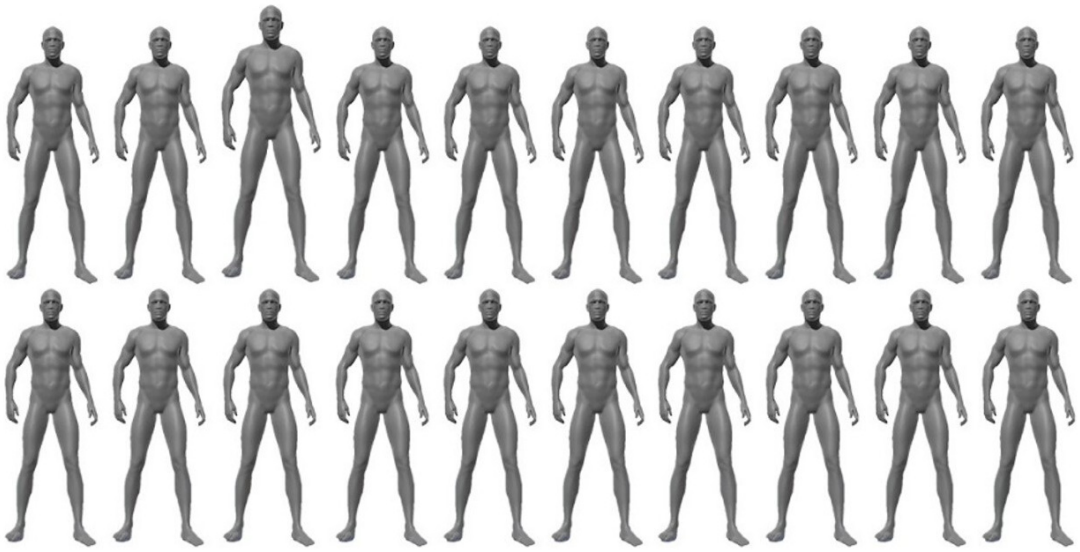


**Figure 6. Stimuli used in Study 7.**

perceiving lower availability of ingroup coalitional support (Prediction 17), and females (Prediction 18), would rate groups of coordinated male-like targets as having higher coalitional fighting capacity.

***Post-task questionnaire***

The post-task questionnaire was the same across Studies 1–7. It included questions about basic demographics, such as biological sex (see formulations on pp. 29–30 in SI4.1 and pp. 24–25 in SI4.2), and measures of individual formidability and perceived coalitional support. To measure individual formidability, we utilized an established measure of self-assessed physical strength (p. 29 in SI4.1), which strongly correlates with objectively measured weightlifting (Sell *et al.*, 2009). This measure is composed of only one item. To reduce measurement error, we derived another item explicitly asking about self-defense capacity (p. 29 in SI4.1). In the pilot data, self-assessed physical strength and defense capacity correlated at $r = .78$. In the main analyses, we combined the two into a single scale. To measure perceived availability of ingroup coalitional support, we derived and refined a 3-item scale tested in the final Pilot 3.16 (in the pilot data, Cronbach's $\alpha = .63$; example item: "If my physical safety was threatened, I could count on many people coming to my defense"). Furthermore, we included an attention check, next to the last question, and two other questions for exploratory analyses about the number of adult male relatives and hostility perceptions (pp. 29–30 in SI4.1). Across P1, P2.1–2.8, and P3.1–3.16, we also asked participants about task difficulty, comfort with installing the Inquisit app, and feedback on the experimental design. Most found the studies engaging and comfortable, offering few critical comments, which we addressed through the 25 pilots. Only a small number correctly guessed the hypotheses, which we clarified during debriefing.

***Examples of online experimental tasks***

Recordings of the experimental tasks can be accessed on the Open Science Framework (OSF) repository associated with the project: https://doi.org/10.17605/OSF.IO/RYV4N.

## Sampling

Studies 1–7 were all single-site, administered online to independent samples of the US Prolific participants in October-December 2024. We conducted the studies sequentially: Studies 1 and 2; Studies 4 and 6; Studies 3 and 5; and Study 7. The participants completed the studies in the setting of their choice. We requested the participants to complete the studies using laptop or PC devices, not phones, because the differences between stimuli categories may be less perceptible on phones with small screens. We selected the US as the study site because Prolific could draw representative samples only in the US and the UK. We selected the US, rather than the UK, because in the former, but not in the latter, we could obtain local ethics approvals without having formal affiliations to local research institutions. Every Prolific participant who was above 18, resided in the US, and was healthy was eligible to participate. The sample was drawn by Prolific platform, following Prolific's probability sampling procedures (see p. 22 in SI4.1 and p. 19 in SI4.2 for the recruitment text). We requested Prolific to generate samples representative of the US population with respect to age, sex, and ethnicity. However, for some studies, namely Studies 3–7, Prolific was unable to generate representative samples within the time promised by the online calculator. Therefore, when the samples of Studies 3–7 were 62%, 62%, 54%, 54%, and 54% completed (respectively), we relaxed the criterion of representativeness with respect to age (the platform contains very few participants in the age category of 55–100). Across Studies 1–7, people who have epilepsy were asked to decline participation upon receiving the invitation from Prolific (for a minority of people with epilepsy, a small risk exists that rapidly presented pictures may trigger photosensitive seizures; Fisher *et al.*, 2022). Samples sizes were determined by power analyses (and estimations of data loss at the pre-processing stage): 800 participants in each Study 1 and 2; 2,300 participants in each Study 3 and 4; 2,800 participants in each Study 5 and 6; and 600 participants in Study 7.

***Online data collection considerations***

The experimental paradigms we use in the present research have mainly been administered in physical labs. Given that we administered these designs online, several notes pertaining to the technical considerations of administration are due. Generally, empirical research demonstrates that results from classical paradigms in cognitive psychology, including those measuring reaction times and those using rapid stimuli presentation, replicate in web-based studies with participants recruited via online research platforms, such as mTurk or Prolific (Crump *et al.*, 2013; Peer *et al.*, 2017). Several studies have shown this to hold true for numerical cognition paradigms specifically (Kochari, 2019). Overall, studies across various areas of experimental psychology/cognitive science demonstrate that the quality of data collected in laboratory and online studies is comparable (Sauter *et al.*, 2020). We selected Prolific, because existing research has validated its data quality, showing comparatively higher rates of attention and comprehension, and lower rates of dishonesty and cheating (Peer *et al.*, 2017; Peer *et al.*, 2022). In our pilots, only about 2% of participants failed at least one attention check. To enhance attentiveness and engagement, we aimed to create tasks that are intrinsically engaging, with participants receiving summaries about their enumeration performance. Qualitative comments, following the experimental tasks indicate that many participants found the tasks engaging. Finally, we used stringent data pre-processing thresholds (see below), including list-wise deletion for participants who failed attention checks.

In online experiments, researchers have little control over the hardware and connection participants use to complete tasks, which can introduce noise both into measurements and stimuli presentation times. We addressed this by using Inquisit software, which we asked the participants to download to their computers. In this way, the app relied on local machine resources, making it independent of the internet speed, reducing the variance in the presentation times and reaction time measurements to single milliseconds. According to our pilot

data, most participants reported that they were comfortable installing the app (p. 5 in SI1.2, p. 4 in SI2.2, and p. 3 in SI3.2).

## Analysis plan

***Studies 1–2***

The dot-probe data was analyzed following three preregistered approaches—standard computation of attentional bias; response-based computation of vigilance scores (Evans & Britton, 2018; Loomis *et al.*, 2023) and mixed-methods analysis of individual-specific attentional bias. The three approaches have their distinct advantages—familiarity and comparability to prior research, internal consistency, and statistical efficiency (respectively)—but their estimates converged in our pilot analyses. The standard approach takes reaction times from congruent trials (RT_congruent) for a particular individual, subtracts them from reaction times from incongruent trials (RT_incongruent) for the same individual, and then averages these RT_differences to produce an aggregate measure of attentional bias toward experimental stimuli. Recent studies found the standard approach to suffer from low internal consistency (Evans & Britton, 2018; Loomis *et al.*, 2023). A computational response-based approach was recently developed to address this issue (Evans & Britton, 2018; Loomis *et al.*, 2023). Following this approach, we computed an average RT from incongruent trials for each participant and then subtracted it from each RT_congruent for the same individuals, eventually computing OLS regressions with such "vigilance" scores as the outcome. Finally, responding to recommendations in methodological literatures (Evans & Britton, 2018; Loomis *et al.*, 2023), we also preregistered a mixed-effects approach. Rather than pre-computing or averaging any elements in the estimation, we used reaction times from all trials as the outcome, a binary variable taking the values of 0 and 1 for, respectively, incongruent and congruent trials as a predictor, and included a participant-specific randomly varying intercept. In the pilots, we did not find significant variance in individual-specific slopes; therefore, in these models we only modeled randomly varying individual-specific intercepts. To formalize, H1 was tested in three linear regression models:

- An OLS regression with RT_difference as an outcome, without predictors or covariates, on the data generated by the trials contrasting coordinated male-like groups vs rotated male-like groups (Prediction 1), then on the data generated by the trials contrasting coordinated male-like groups vs large coordinated male-like groups (Prediction 2), and lastly on the data generated by the trials contrasting rotated male-like groups vs large coordinated male-like groups (Prediction 3).

- An OLS regression with vigilance scores as an outcome, without predictors or covariates, on the same data (Predictions 1–3).

- A mixed-effects linear regression with RT as an outcome, a binary indicator of congruence/incongruence as a predictor, a randomly varying individual-specific intercept, without any other covariates, on the same data (Predictions 1–3).

***Studies 3–4***

Studies 3–4 focused on estimation accuracy. We analyzed two outcomes: a continuous value indicating the size of estimation error (actual target N – estimated target N) and a binary indicator of correct = 1 vs incorrect = 0. The continuous outcome was analyzed in a mixed-effects model with an identity link function and the binary outcome in a mixed-effects model with a logit link function. Randomly varying individual-specific intercepts were included in both models. In the pilots, we did not find consistent variation in regression slopes; therefore, we did not include randomly varying slopes for individuals. No other covariates were included in the specification. More formally, the hypothesis that people enumerate potentially threatening coalitions more accurately (H2) were tested in:

- A linear (identity link function) mixed-effects regression model with a randomly varying individual-specific intercept, a (continuous) estimation error as the outcome variable, and a binary predictor taking the value of 1 for coordinated male-like targets and 0 for rotated male-like targets (Prediction 4), and then in a separate identical model but with a binary predictor taking the value of 1 for coordinated male-like targets and 0 for coordinated female-like targets (Prediction 5).

- A generalized linear (logit link function) mixed-effects regression model with a randomly varying individual-specific intercept, a binary indicator of correct/incorrect response as the outcome variable, and a binary predictor taking the value of 1 for coordinated male-like targets and 0 for rotated male-like targets (Prediction 4), and then in a separate identical model but with a binary predictor taking the value of 1 for coordinated male-like targets and 0 for coordinated female-like targets (Prediction 5).

The hypothesis that males determine the size of potentially threatening coalitions more accurately than females (H9) was tested in:

- A generalized linear (logit link function) mixed-effects regression model with a randomly varying individual-specific intercept, a binary indicator of correct/incorrect response as the outcome variable, and a dummy for participant sex as the predictor, using only the trials with coordinated male-like targets (Prediction 6).

The hypotheses that females (H11), less formidable individuals (H13), and individuals with limited access to coalitional support (H14) tend to overestimate the size of potentially threatening coalitions were tested in three mixed-effects logit models using a measure of positive error as the outcome. This measure is a dummy variable taking the value 1 if the participant's estimation was higher than the actual number of targets

in a trial, and 0 if the participant's estimation was correct or lower than the actual number of targets in the trial. To test H11, H13, and H14 we ran (respectively):

- A generalized linear (logit link function) mixed-effects regression model with a randomly varying individual-specific intercept, the positive error dummy as the outcome variable, and a dummy for participant sex as the predictor (Prediction 7).

- A generalized linear (logit link function) mixed-effects regression model with a randomly varying individual-specific intercept, the positive error dummy as the outcome variable, and the 2-item participant formidability scale as the predictor (Prediction 8).

- A generalized linear (logit link function) mixed-effects regression model with a randomly varying individual-specific intercept, the positive error dummy as the outcome variable, and the 3-item coalitional support scale as the predictor (Prediction 9).

***Studies 5–6***

Studies 5–6 focused on counting rapidity, i.e., the time between the onset of the stimuli and the counting of targets as indicated with pressing of the keyboard keys. Analogously to Studies 3–4, we relied on mixed-effects models with individual-specific randomly varying intercepts. No individual-level covariates were included in the specification. More formally, the hypothesis that people enumerate potentially threatening coalitions more rapidly (H3) was tested in:

- A linear (identity link function) mixed-effects regression model with a randomly varying individual-specific intercept, counting latency as the outcome variable, and a binary predictor taking the value of 1 for coordinated male-like targets and 0 for rotated male-like targets (Prediction 10), and then in a separate identical model but with a binary predictor taking the value of 1 for coordinated male-like targets and 0 for coordinated female-like targets (Prediction 11).

The hypothesis that males determine the size of threatening coalitions more rapidly than females (H10) were tested in:

- A linear (identity link function) mixed-effects regression model with a randomly varying individual-specific intercept, counting latency as the outcome variable, and a dummy for participant sex as the predictor, using only the trials with coordinated male-like targets (Prediction 12).

***Study 7***

The hypothesis that subjective coalitional formidability assessments are a function of the N of targets (H4) and the formidability of individual members of the coalition (H5) was tested in (respectively):

- A linear (identity link function) mixed-effects regression with a randomly varying individual-specific intercept, with the subjective coalitional formidability ratings as the outcome variable, and the N of targets in each trial, ranging from 4 to 10, as the predictor (Prediction 13).

- A linear (identity link function) mixed-effects regression with a randomly varying individual-specific intercept, with the subjective coalitional formidability ratings as the outcome variable, and a binary predictor indicating the presence (1) or absence (0) of one larger target in the coalition in the trial (Prediction 14).

The hypothesis that the effect of a more formidable individual in the coalition on subjective formidability assessment diminishes with group size (H6) was tested in:

- A linear (identity link function) mixed-effects regression with a randomly varying individual-specific intercept, with the subjective coalitional formidability ratings as the outcome variable, and a multiplicative interaction term between the N of targets and the dummy indicating presence of one larger target (Prediction 15).

The hypothesis that subjective formidability assessments of potentially threatening coalitions are a negative function of one's own formidability (H7) was tested in:

- A linear (identity link function) mixed-effects regression with a randomly varying individual-specific intercept, with the subjective coalitional formidability ratings as the outcome variable, and the 2-item participant formidability scale as the predictor (Prediction 16).

The hypothesis that subjective formidability assessments of potentially threatening coalitions are a negative function of perceived availability of ingroup coalitional support (H8) was tested in:

- A linear (identity link function) mixed-effects regression with a randomly varying individual-specific intercept, with the subjective coalitional formidability ratings as the outcome variable, and the 3-item coalitional support scale as the predictor (Prediction 17).

The hypothesis that females, compared to males, subjectively assess potentially threatening coalitions as more formidable (H12) was tested in:

- A linear (identity link function) mixed-effects regression with a randomly varying individual-specific intercept, with the subjective coalitional formidability ratings as the outcome variable, and a dummy for participant sex as the predictor (Prediction 18).

No individual-level covariates were included in any of the specifications of Study 7.

***Data pre-processing***

Following research ethics principles, we excluded all participants who decided to withdraw their data at the end of the study. In some of the pilots we found duplicate observations, which were also deleted prior to analysis. We also removed all data generated via iOS or Android devices. Participants who failed attention checks were also removed. For post-task questionnaire data, we removed observations with "prefer not to state". For the dot-probe analyses (Studies 1 and 2), we removed untypically slow and fast reaction times. We preregistered

900ms and 200ms as the cut-offs, roughly corresponding to 1st and 99th percentiles in the pilot data. This restricted our data to high quality responses provided by engaged participants, as contrasted to careless, inattentive, and random responders or "straight liners". Only correct responses, indicated by N or M clicks corresponding to X appearing on, respectively, left and right, were analyzed (only 0.13% of responses were incorrect in the pilot data). For the counting rapidity analyses (Studies 5 and 6), we used a listwise-deletion for trial-specific latencies above 4000ms. Approximately 95% of observations fall below this threshold in the pilot data. A 95th percentile trim is considerable, and much larger than the 99th percentile trim used in the dot-probe trials. This stringent trim was informed by extant research showing that it takes 40–200ms/item to enumerate items within the subitizing range (1-4 items) and 250–350ms/item when there are more than 4 items (Trick & Pylyshyn, 1994). Hence, even if we assume that no subitizing appears for the subsets of 8 items (i.e., all 8 items are counted, with none subitized), and use the upper estimated bound, the expected counting latency for 8 items should be less than 2800ms, making 4000ms threshold liberal. No other data pre-processing measures were taken and no transformations of scores were used.

### Deviations from preregistration

In SI5, we report deviations from the preregistered data collection and analysis plan, such as those resulting from typographical errors in the original protocol, reductions in sample size due to, for example, duplicate responses from Prolific participants, and incorrect model specifications.

## Results

### Study 1

***Summary***

Study 1 revealed an attentional bias toward coordinated male-like groups when contrasted with rotated male-like groups that are otherwise identical, supporting Prediction 1 (Figure 7). It also found biases toward larger coordinated male-like groups when contrasted with smaller coordinated and rotated male-like groups, supporting Predictions 2 and 3 (Figure 7). Notably, the rapid and spontaneous attentional shifts observed in the dot-probe task closely aligned with participants' explicit ratings of threat and coalitionality, both of which increased consistently across the same group contrasts—from rotated to coordinated to larger coordinated male-like groups.

***Details***

A total of 774 participants completed Study 1[2]. Among these, 30 eventually requested to delete their data, 41 used Android or iOS devices, and 12 failed the attention check. Study 1 thus retained 691 participants. Consistent with pilots, participants rated the threat of rotated male-like groups, coordinated male-like groups, and larger coordinated male-like groups in increasing order: ***M* = 1.94**, 95% CI = [1.86, 2.02]; ***M* = 2.45**, 95% CI = [2.35, 2.55]; and ***M* = 3.20**, 95% CI = [3.08, 3.32]. Coalitionality ratings followed the same pattern: ***M* = 3.14**, 95% CI = [3.02, 3.26]; ***M* = 4.86**, 95% CI = [4.75, 4.98]; and ***M* = 5.68**, 95% CI = [5.58, 5.78]. Threat and coalitionality ratings correlated at $r$ = .30, .30, and .39, respectively.

The standard computational approach revealed the following **automatic attentional biases**: **4.0 ms** ($p$ = 0.049, 95% CI = [0.0, 8.1]) toward coordinated male-like groups when contrasted with rotated male-like groups; **6.7 ms** ($p$ = 0.001, 95% CI = [2.8, 10.6]) toward larger coordinated male-like groups when contrasted with smaller coordinated male-like groups; and **10.5 ms** ($p$ < 0.001, 95% CI = [6.4, 14.6]) toward larger coordinated male-like groups when contrasted with smaller rotated male-like groups. A computational response-based approach revealed similar estimates: **3.6 ms** ($p$ = 0.023, 95% CI = [0.5, 6.6]), **6.8 ms** ($p$ < 0.001, 95% CI = [3.7, 10.0]), and **10.2 ms** ($p$ < 0.001, 95% CI = [7.0, 13.4]); and so did the linear mixed-effects regression models: **4.0 ms** ($p$ = 0.048, 95% CI = [0.0, 8.0]), **7.1 ms** ($p$ = 0.001, 95% CI = [3.1, 11.1]), and **11.0 ms** ($p$ < 0.001, 95% CI = [6.9, 15.0]).

### Study 2

***Summary***

Study 2 evaluated an alternative explanation for the observed bias toward coordinated male-like groups (compared to rotated male-like groups), specifically the possibility that it is the coordination or similarity of targets per se that attracts attention. The results did not support this explanation: coordinated female-like targets, when contrasted with rotated female-like targets, did not elicit increased spontaneous attention (Figure 7). Taken together with the findings from Study 1, Study 2 revealed that it is not coordination alone, but coordination among males, that draws attention. Additionally, as outlined in our preregistration, we expected that coordinated male-like groups would not attract significantly more attention than coordinated female-like groups, as the attractiveness of the targets might suppress attention due to perceived threat. The results of Study 2 were consistent with this expectation, although the estimates indicated a consistent trend toward greater attention to coordinated male-like groups and some coefficients approached statistical significance (Figure 7). However, we predicted that stronger cues of coalitionality and threat—operationalized as eight coordinated male-like targets—would override the effects of attractiveness. This prediction was supported across all models in Study 2 (Figure 7).

***Details***

A total of 785 participants completed Study 2[3]. Among these, 34 eventually requested to delete their data, 37 used Android or iOS devices, and 18 failed the attention check. Another 6 completed only part of the study, and since the attention check was at the end of the study, they were also excluded. Study 2 thus retained 690 participants.

Consistent with pilots, participants rated the threat of rotated female-like groups, coordinated female-like groups, coordinated

[2] Prolific supplied data with 800 entries; however, 26 of these were duplicates (i.e., provided by the same participants).

[3] Once again, Prolific supplied data with 800 entries but 15 of these were duplicates.

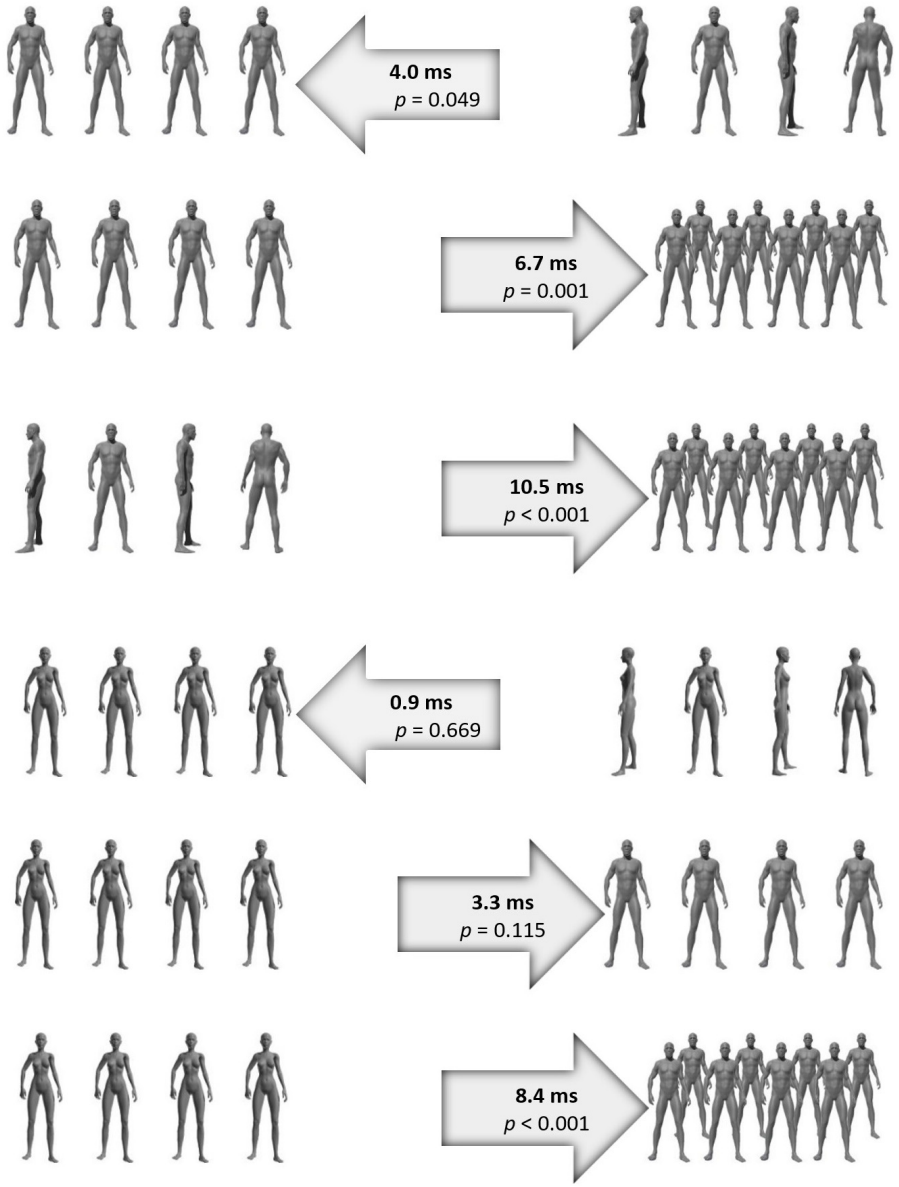



**Figure 7. Study 1 and 2 results: Automatic attention bias toward "coalitions" presented in pairs.**

male-like groups, and larger coordinated male-like groups in increasing order: ***M* = 1.39**, 95% CI = [1.33, 1.45]; ***M* = 1.66**, 95% CI = [1.59, 1.74]; ***M* = 2.39**, 95% CI = [2.29, 2.49]; and ***M* = 3.12**, 95% CI = [3.00, 3.24]. Coalitionality ratings followed the same pattern: ***M* = 3.07**, 95% CI = [2.96, 3.19]; ***M* = 4.70**, 95% CI = [4.58, 4.81]; ***M* = 4.85**, 95% CI = [4.75, 4.97]; and ***M* = 5.59**, 95% CI = [5.49, 5.70]. Threat and coalitionality ratings correlated, more so for male-like targets ($r$ = .28 and .36) than for female-like targets ($r$ = .19 and .20). Subsequently, corresponding to Study 1, threat and coalitionality cues attracted automatic attention.

The standard computational approach revealed no significant attentional bias toward coordinated female-like groups, when contrasted to rotated female-like groups, **0.9ms** ($p$ = 0.669, 95% CI = [-3.3, 5.1]). Attentional bias toward coordinated male-like groups, when contrasted to coordinated female-like groups, was larger but also insignificant, **3.3ms** ($p$ = 0.115, 95% CI = [-0.8, 7.3]). The latter bias increased when the N of male-like targets was doubled, **8.4ms** ($p$ < 0.001, 95% CI = [4.0, 12.8]). A computational response-based approach revealed consistent estimates: **2.2 ms** ($p$ = 0.199, 95% CI = [-1.2, 5.5]), **3.0 ms** ($p$ = 0.067, 95% CI = [-0.2, 6.3]), and **7.7 ms** ($p$ < 0.001,

95% CI = [4.2, 11.]); and so did linear mixed-effects regression models: **1.7 ms** ($p$ = 0.424, 95% CI = [-2.5, 6.0]), **3.0 ms** ($p$ = 0.143, 95% CI = [-1.0, 7.1]), and **8.1 ms** ($p$ < 0.001, 95% CI = [3.8, 12.4]).

## Study 3

***Summary***

Study 3 revealed that people made fewer mistakes estimating coordinated male-like targets than rotated ones, supporting Prediction 4 (Figure 8). It also found that male participants were more accurate than female when estimating coordinated male-like targets, supporting Prediction 6. Furthermore, female participants and those who saw themselves as less formidable tended to overestimate the targets, which supports Predictions 7 and 8. In contrast, perceived availability of ingroup coalitional support was not significantly associated with overestimation of coordinated male-like targets, failing to support Prediction 9.

***Details***

A total of 2,266 participants completed Study 3[4]. Among these, 109 eventually requested to delete their data, 115 used Android or iOS devices, and 39 failed the attention check. Another 7 completed only part of the study and so were also excluded. Study 3 thus retained 1996 participants.

Linear mixed-effects regressions revealed that coordinated male-like targets were estimated with a lower error (**average predicted error = -0.036**, 95% CI = [-0.045, -0.027]) compared to rotated male-like targets (**-0.057**, 95% CI = [-0.067, -0.048]) and dots (**-0.095**, 95% CI = [-0.105, -0.086]). The differences in estimation errors were significant for both contrasts:

[4] Once again, Prolific supplied data with 2300 entries but 34 of these were duplicates

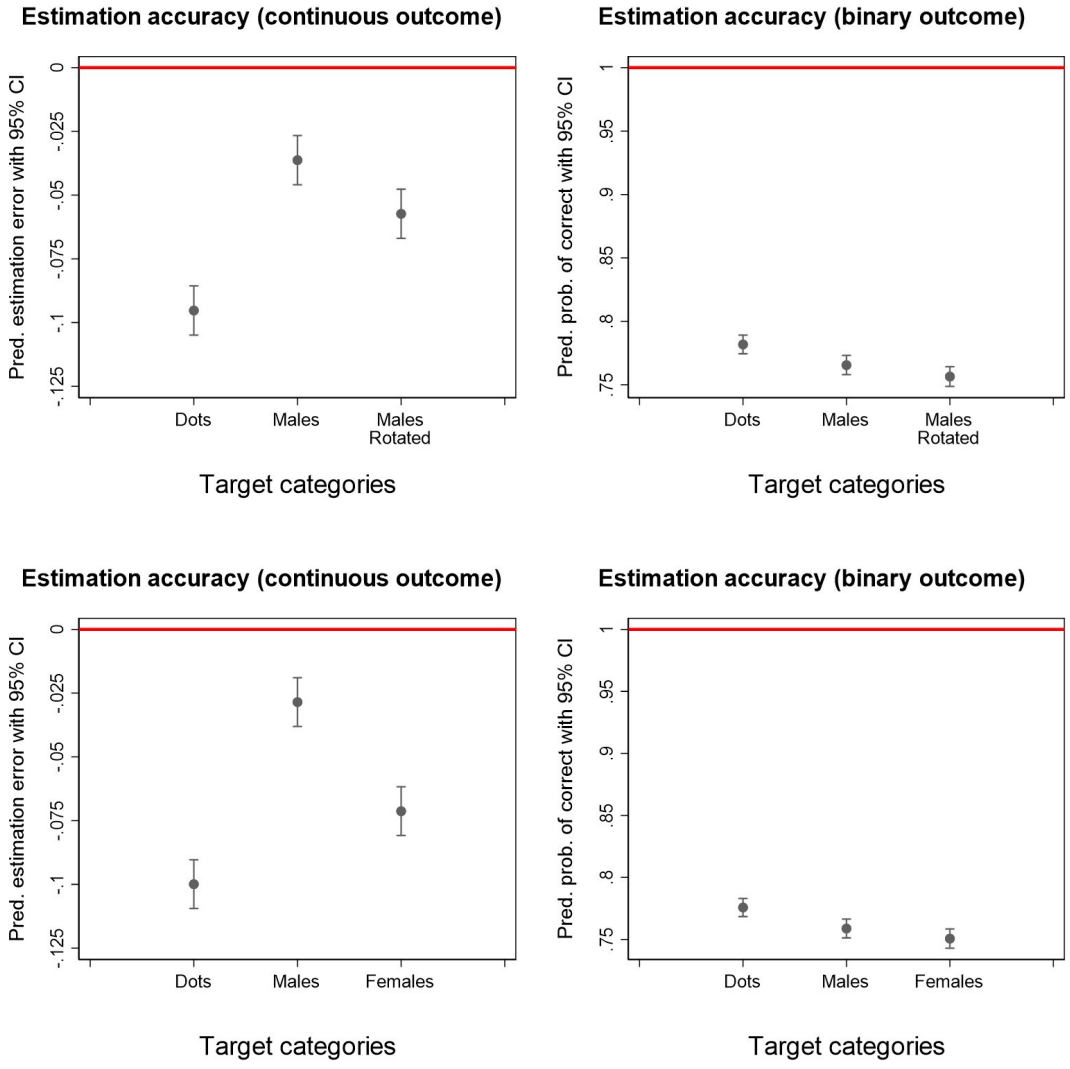



**Figure 8. Study 3 and 4 results: Estimation accuracy for dots and "coalitions".**

coordinated male-like targets vs. dots (**-0.059**, $p < 0.001$, 95% CI = [-0.068, -0.049]) and coordinated male-like targets vs. rotated male-like targets (**-0.021**, $p < 0.001$, 95% CI = [-0.031, -0.012]). Consistent with pilots, non-preregistered tests with target *N* limited to 8 revealed even larger differences in estimation errors (respectively): **-0.074**, $p < 0.001$, 95% CI = [-0.093, -0.055]; **-0.036**, $p < 0.001$, 95% CI = [-0.055, -0.017].

Logit mixed-effects regressions revealed that the rate of correct responses was highest for dots (**marginal predicted mean × 100 = 78.2%** [77.4%, 78.9%]), intermediate for coordinated male-like targets (**76.6%**, 95% CI = [75.8%, 77.4%]), and lowest for rotated male-like targets (**75.6%**, 95% CI = [74.9%, 76.4%]). The difference between coordinated male-like targets and non-coordinated ones was significant: **logit = -0.054**, $p = 0.015$, 95% CI = [-0.099, -0.011].

Analogous analyses revealed that the rate of correct responses for coordinated male-like targets was higher among male participants (**marginal predicted mean × 100** = 77.6%, 95% CI = [76.6%, 78.6%]) compared to female participants (75.2%, 95% CI = [74.0%, 76.3%]). The difference was significant: **logit** = 0.147, $p = 0.002$, 95% CI = [0.054, 0.241].

In turn, logit mixed-effects regressions revealed that the rate of overestimation errors for coordinated male-like targets was higher among female participants (**marginal predicted mean × 100** = 11.6%, 95% CI = [10.8%, 12.5%]) compared to male participants (**9.2%**, 95% CI = [8.6%, 9.6%]). Female sex was a significant predictor of overestimation errors: **logit** = 0.282, $p < 0.001$, 95% CI = [0.159, 0.404,]. Likewise, the 2-item participant formidability scale, derived from questions about self-perceived strength and defense capacity ($r = .72$), significantly predicted overestimation errors, with lower scores associated with higher rates of overestimation: **logit** = -0.040, $p = 0.005$, 95% CI = [-0.069, -0.012]. By contrast, in a non-preregistered analysis, neither sex nor participant formidability significantly predicted overestimation errors for dots, respectively: **logit** = 0.114, $p = 0.110$, 95% CI = [-0.026, 0.253] and **logit** = -0.005, $p = 0.740$, 95% CI = [-0.037, 0.026]. Finally, the 3-item scale of perceived coalitional support ($\alpha = 0.74$) did not significantly predict overestimation errors: **logit** = 0.017, $p = 0.502$, 95% CI = [-0.033, 0.067].

## Study 4

### *Summary*

Study 4 was designed to evaluate an alternative explanation for Prediction 4—namely, that the improved enumeration of coordinated male-like targets may be driven by a similarity grouping effect. To test this, we specified Prediction 5, in which both stimulus categories (male-like and female-like) included aligned targets. If coordinated male-like targets were still enumerated more accurately than coordinated female-like targets, this would suggest a category-based effect beyond mere similarity. Study 4 supported this prediction (Figure 8). Moreover, Study 4 replicated the Study 3 findings pertaining to Predictions 6–8, demonstrating in an independent sample that male participants produced lower estimation errors than female participants, and that female participants and those with lower self-assessed formidability tended to overestimate the number of male-like targets.

### *Details*

A total of 2,267 participants completed Study 4[5]. Among these, 79 eventually requested to delete their data, 121 used Android or iOS devices, and 35 failed the attention check. Another 5 completed only part of the study and so were also excluded. Study 4 thus retained 2027 participants.

Linear mixed-effects regressions revealed that coordinated male-like targets were estimated with a lower error (**average predicted error = -0.028**, 95% CI = [-0.038, -0.019]) compared to coordinated female-like targets (**-0.071**, 95% CI = [-0.081, -0.019]) and dots (**-0.100**, 95% CI = [-0.109, -0.090]). The differences in estimation errors were significant for both contrasts: coordinated male-like targets vs. dots (**-0.071**, $p < 0.001$, 95% CI = [-0.081, -0.062]) and coordinated male-like targets vs. coordinated female-like targets (**-0.043**, $p < 0.001$, 95% CI = [-0.052, -0.033]). Consistent with pilots, analogous (non-preregistered) tests with target *N* limited to 8 revealed even larger differences in estimation errors, respectively: **-0.082**, $p < 0.001$, 95% CI = [-0.101, -0.063]; **-0.056**, $p < 0.001$, 95% CI = [-0.075, -0.037].

Logit mixed-effects regressions revealed that the rate of correct responses was highest for dots (**marginal predicted mean × 100 = 77.6%** [76.8%, 78.3%]), intermediate for coordinated male-like targets (**75.9%**, 95% CI = [75.1%, 76.6%]), and lowest for coordinated female-like targets (**75.1%**, 95% CI = [74.3%, 75.8%]). The difference between coordinated male-like targets and non-coordinated ones was significant: **logit = -0.048**, $p = 0.029$, 95% CI = [-0.092, -0.005]. Analogous analyses revealed that the rate of correct responses for coordinated male-like targets was higher among male participants (**marginal predicted mean × 100** = 77.8%, 95% CI = [76.7%, 78.8%]) compared to female participants (74.3%, 95% CI = [73.2%, 75.3%]). The difference was significant: **logit** = 0.212, $p < 0.001$, 95% CI = [0.121, 0.302].

In turn, logit mixed-effects regressions revealed that the rate of overestimation errors for coordinated male-like targets was higher among female participants (**marginal predicted mean × 100 = 12.3%**, 95% CI = [11.4%, 13.1%]) compared to male participants (**9.4%**, 95% CI = [8.7%, 10.2%]). The female sex was a significant predictor of overestimation errors: **logit** = 0.320, $p < 0.001$, 95% CI = [0.196, 0.444]. Likewise, the 2-item formidability scale, derived from questions about self-perceived strength and defense capacity ($r = .74$), significantly predicted overestimation errors, with lower scores associated with higher rates of overestimation: **logit** = -0.037, $p = 0.013$, 95% CI = [-0.008, -0.066]. A non-preregistered analysis revealed that female sex and low participant formidability also significantly predicted overestimation errors for dots, respectively: **logit** = 0.178, $p = 0.010$, 95%

[5] Prolific supplied data with 2300 entries but 33 of these were duplicates.

CI = [0.041, 0.315] and **logit** = -0.043, $p$ = 0.010, 95% CI = [-0.074, -0.010]. However, an analysis of a multiplicative interaction term, female sex*target category, revealed that female participants produced higher overestimation error rates for coordinated male-like targets (**marginal predicted mean × 100 = 12.3%**, 95% CI = [11.5%, 13.1%]) than for dots (**7.3%**, 95% CI = [6.7%, 7.9%]). The interaction term was significant at $p$ = 0.019. By contrast, an analogous interaction term for participant formidability*target category was insignificant, $p$ = 0.893. Finally, the 3-item scale of perceived coalitional support ($\alpha$ = 0.74) did not significantly predict overestimation errors: **logit** = -0.012, $p$ = 0.631, 95% CI = [-0.062, 0.038].

## Study 5

***Summary***

Study 5 revealed that participants counted coordinated male-like targets faster than rotated male-like targets, supporting Prediction 10. It also demonstrated that male participants, compared to female, counted faster coordinated male-like targets, supporting Prediction 12.

***Details***

Power analyses for Studies 5 and 6 indicated that 2,800 participants were required; however, Prolific imposed a maximum limit of 2,500 participants for representative samples. A total of 2,469 participants completed Study 5[6]. Among these, 105 eventually requested to delete their data, 146 used Android or iOS devices, and 39 failed the attention check. Another 39 completed only part of the study and so were also excluded. Study 5 thus retained 2140 participants.

Linear mixed-effects regressions revealed that coordinated male-like targets were counted faster (**average predicted latency = 1957.1ms**, 95% CI = [1933.5, 1980.7]) compared to rotated male-like targets (**1977.1ms**, 95% CI = [1953.4, 2000.7]), and slower compared to dots (**1935.5ms**, 95% CI = [1911.8, 1959.1]). The differences in counting latencies were significant for both the coordinated male-vs-rotated male contrast (**20.0ms**, $p$ < 0.001, 95% CI = [9.7, 30.3]) and for the coordinated male-vs-dot contrast (**-21.6ms**, $p$ < 0.001, 95% CI = [-31.9, -11.34]). In contrast to pilots, non-preregistered tests with target $N$ limited to 8 revealed that male-like targets were not counted faster (**average predicted latency = 2201.8ms**, 95% CI = [2175.1, 2228.6]) compared to rotated male-like targets (**2200.0ms**, 95% CI = [2173.2, 2226.7]) and slower than dots (**2168.7ms**, 95% CI = [2142.0, 2195.4]).

In turn, analyses revealed that counting latencies for coordinated male-like targets was lower among male participants (**average predicted latency = 1863.1ms**, 95% CI = [1832.4, 1893.9]) compared to female participants (**2056.3ms**, 95% CI = [2022.0, 2090.6]). The male sex was a significant predictor of counting latency (**-193.1ms**, $p$ < 0.001, 95% CI = [-239.2, -147.1]). A non-preregistered test revealed that male sex also significantly predicted counting latencies for dots, although the estimate was smaller (**-179.8ms**, $p$ < 0.001, 95% CI = [-226.8, -132.8]).

## Study 6

***Summary***

Analogously to Study 4, Study 6 was designed to evaluate an alternative explanation for Prediction 10—namely, that the improved counting latency of coordinated male-like targets may be driven by a similarity grouping effect. To test this, we specified Prediction 11, in which both stimulus categories (male-like and female-like) included aligned targets. If coordinated male-like targets were still enumerated faster than coordinated female-like targets, this would suggest a category-based effect beyond mere similarity. Study 6 only supported this prediction when target N was limited to 8. We return to this finding in Discussion. Additionally, Study 6 replicated the Study 5 findings pertaining to Prediction 12, demonstrating in an independent sample that male participants, compared to female, counted coordinated male-like targets faster.

***Details***

A total of 2,484 participants completed Study 6[7]. Among these, 77 eventually requested to delete their data, 137 used Android or iOS devices, and 48 failed the attention check. Another 18 completed only part of the study and so were also excluded. Study 6 thus retained 2204 participants.

Linear mixed-effects regressions revealed that coordinated male-like targets were counted negligibly faster (**average predicted latency = 1971.5ms**, 95% CI = [1949.1, 1994.0]) compared to coordinated female-like targets (**1972.6ms**, 95% CI = [1950.2, 1995.1]), and considerably slower compared to dots (**1934.7ms**, 95% CI = [1912.3, 1957.2]). The differences in counting latencies were not significant for the male-vs-female contrast (**1.1ms**, $p$ = 0.830, 95% CI = [-9.0, 11.2]) and significant for the male-vs-dot contrast (**-36.8**, $p$ < 0.001, 95% CI = [-46.85, -26.74]).

Consistent with pilots, non-preregistered tests with target $N$ limited to 8 revealed that male-like targets were counted faster (**average predicted latency = 2201.1ms**, 95% CI = [2175.6, 2226.4]) compared to coordinated female-like targets (**2223.4ms**, 95% CI = [2198.1, 2248.8]) and slower than dots (**2163.9ms**, 95% CI = [2138.6, 2189.3]). The differences in counting latencies were significant for both contrasts: coordinated male-like targets vs. female-like (**22.5ms**, $p$ = 0.008, 95% CI = [5.73, 39.2]) and coordinated male-like targets vs. dots (**-37.1ms**, $p$ < 0.001, 95% CI = [-53.8, -20.37]).

Consistent with Study 5, analyses revealed that counting latencies for coordinated male-like targets was lower among male participants (**average predicted latency = 1911.8ms**, 95%

[6] Prolific supplied data from 2500 participants but 31 of these did not complete a single trial and so were removed prior to the data pre-processing steps.

[7] Prolific supplied data from 2501 participants but 17 of these did not complete a single trial and so were removed prior to the data pre-processing steps.

CI = [1881.6, 1942.0]) compared to female participants (**2011.1ms**, 95% CI = [1980.0, 2041.9]) (Prediction 12). The male sex was a significant predictor of counting latency (**-103.3ms**, $p < 0.001$, 95% CI = [-147.9, -58.9]). A non-preregistered test revealed that male sex also significantly predicted counting latencies for dots, although, the coefficient was smaller (**-94.1**, $p < 0.001$, 95% CI = [-138.0, -50.19]).

### Study 7

***Summary***

Study 7 supported Prediction 13, showing that subjective assessments of coalitional formidability increased as a function of the number of individuals in the coalition. Prediction 14 was also supported: the presence of one larger individual within the coalition led to higher formidability ratings. Consistent with Prediction 15, this effect of the more formidable individual diminished as the overall coalition size increased. In contrast, neither Prediction 16 nor Prediction 17 were supported: self-reported individual formidability and perceived availability of ingroup coalitional support did not significantly predict subjective assessments of the target coalitions. Prediction 18 was supported: female participants assessed target coalitions as more formidable than did male participants.

***Details***

A total of 603 participants completed Study 7. Among these, 44 eventually requested to delete their data, 27 used Android or iOS devices, and 19 failed the attention check. Another 3 completed only part of the study and so were also excluded. Study 7 thus retained 510 participants.

Linear mixed-effects regressions revealed that the number of targets in the stimulus array had a significant positive effect on subjective coalitional formidability assessments. Each additional target increased the **predicted coalitional formidability rating by 0.613 points** ($p < 0.001$, 95% CI = [0.599, 0.627]). Likewise, the presence of one larger target in the array significantly increased subjective formidability assessments, with groups containing a larger individual receiving a predicted **0.088-point** higher rating ($p = 0.037$, 95% CI = [0.006, 0.170]). Including a multiplicative interaction between the number of targets and the presence of a larger target revealed a significant negative effect (**b = -0.066**, $p < 0.001$, 95% CI = [-0.094, -0.037]), indicating that the formidability-enhancing effect of one larger target was attenuated with increasing N of targets. The effect of one large target on coalitional formidability ratings disappeared once that target was part of coalitions larger than 6 (Figure 9).

In turn, participant sex significantly predicted subjective coalitional formidability ratings, with female participants rating stimulus arrays as more formidable than male participants (**b = -0.308**, $p = 0.002$, 95% CI = [-0.499, -0.116]). Notably, non-preregistered analyses revealed that the sex difference in formidability ratings decreased as the number of targets increased, as indicated by a significant positive interaction between sex and number of targets in the array (**b = 0.047**, $p = 0.002$, 95% CI = [0.018, 0.076]). Finally, subjective coalitional formidability ratings were not significantly predicted by participants' own individual formidability (**b = -0.009**, p = 0.692, 95% CI = [-0.054, 0.036]) or by their perceived availability of ingroup coalitional support (**b = -0.022**, p = 0.592, 95% CI = [-0.102, 0.058].

## Discussion

The observed results, summarized in Table 1, suggest that human minds are equipped with mechanisms for detecting, enumerating, and assessing potentially threatening coalitions. Participants automatically attended to coordinated male-like groups—stimuli that evolutionarily likely cued for threat of a coalitional attack. These coalitions were enumerated more accurately and rapidly than control stimuli, particularly by male participants. Female participants, and those lower in self-assessed formidability, tended to overestimate the size of target coalitions. Larger coalitions and those containing a standout individual were rated as more formidable, although the

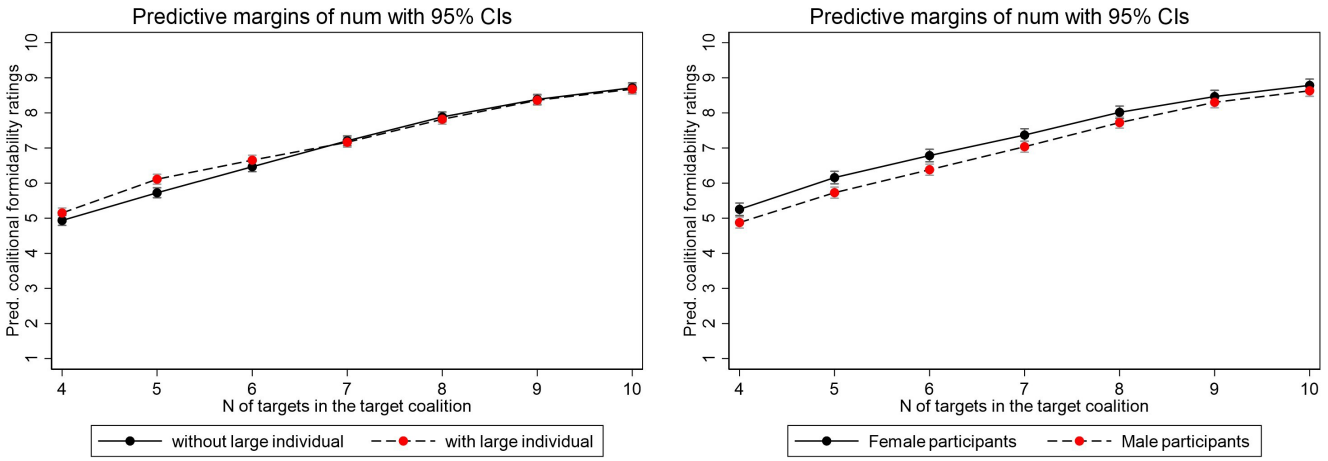



**Figure 9. Study 7 results: Formidability assessments of "coalitions".**

**Table 1. Summary of results: Studies 1–7.**

| Hypothesis | Supported | Comment | Study |
|---|---|---|---|
| People attend, automatically, to potentially threatening coalitions (H1) | Yes | An attentional bias toward coordinated male-like groups when contrasted with rotated male-like groups was identified (Prediction 1). Also, attentional biases toward larger coordinated male-like groups when contrasted with smaller coordinated and rotated male-like groups were identified (Predictions 2 and 3). | 1 and 2 |
| People enumerate, accurately (H2), members of potentially threatening coalitions | Yes | Coordinated male-like targets were estimated with lower error compared to rotated male-like targets (Prediction 4). Coordinated male-like targets were also estimated with lower error compared to coordinated female-like targets (Prediction 5), indicating that the accuracy of enumeration depends on the category of the targets and not just their alignment. | 3 and 4 |
| People enumerate, rapidly (H3), members of potentially threatening coalitions | Partly | Coordinated male-like targets were counted faster compared to rotated male-like targets (Prediction 10). Coordinated male-like targets were not counted faster compared to coordinated female-like targets (Prediction 11). | 5 and 6 |
| Formidability assessments of potentially threatening coalitions are a positive function of their numerical size (H4) | Yes | The number of targets in the stimulus array had a significant positive effect on subjective coalitional formidability assessments, with each additional target increasing the predicted formidability rating (Prediction 13). | 7 |
| Formidability assessments of potentially threatening coalitions are a positive function of the formidability of their individual members (H5) | Yes | The presence of one larger target in the array significantly increased subjective coalitional formidability assessments (Prediction 14). | 7 |
| ...but this function diminishes with the numerical size of the outgroup coalitions (H6) | Yes | A significant negative interaction was identified, indicating that the formidability-enhancing effect of one larger target was attenuated with increasing N of targets (Prediction 15). | 7 |
| Formidability assessments of potentially threatening coalitions are a negative function of one's own formidability (H7) | No | Subjective coalitional formidability ratings were not significantly predicted by participants' individual formidability scores (Prediction 16). | 7 |
| Formidability assessments of potentially threatening coalitions are a negative function of perceived availability of ingroup coalitional support (H8) | No | A scale of perceived coalitional support did not significantly predict overestimation errors for coordinated male-like targets (Prediction 17). | 7 |
| Males, compared to females, determine the size of potentially threatening coalitions more accurately (H9) | Yes | Male participants, compared to female, produced smaller estimation errors for coordinated male-like targets (Prediction 6). In turn, the rate of correct responses for coordinated male-like targets was higher among male participants compared to female participants. | 3 and 4 |
| Males, compared to females, determine the size of potentially threatening coalitions more rapidly (H10) | Yes | Counting latencies for coordinated male-like targets were lower among male participants compared to female participants (Prediction 12). Note that a comparable difference was also observed for dot targets. | 5 and 6 |
| When females make errors in these estimates, they tend to overestimate more so than males (H11) | Yes | The rate of overestimation errors for coordinated male-like targets was higher among female participants compared to male participants (Prediction 7). | 3 and 4 |
| Females, compared to males, subjectively assess potentially threatening coalitions as more formidable (H12) | Yes | Participant sex significantly predicted subjective coalitional formidability ratings, with female participants rating stimulus arrays as more formidable than male participants (Prediction 18). However, the sex difference in formidability ratings decreased as the number of targets increased. | 7 |
| Less formidable individuals tend to overestimate the size of potentially threatening coalitions (H13) | Yes | A participant's formidability scale significantly predicted overestimation for coordinated male-like targets, with lower scores associated with higher rates of overestimation (Prediction 8). | 3 and 4 |
| Individuals with limited access to ingroup coalitional support tend to overestimate the size of potentially threatening coalitions (H14) | No | A scale of perceived coalitional support did not significantly predict overestimation errors for coordinated male-like targets (Prediction 9). | 3 and 4 |

effect of the standout diminished with increasing group size. The sex differences observed across tasks suggest that males have evolved more refined systems for assessing coalitional threats, while females may have evolved cautious perceptual biases. Males, compared to females, consistently demonstrated greater speed and accuracy in coalition enumeration tasks. Females, in contrast, perceived coalitions as more formidable and were more likely to overestimate their size. Importantly, the sex differences in formidability ratings diminished as coalition size increased, suggesting that larger coalitions render sex-linked perceptual biases less relevant.

Contrary to expectations, self-assessed individual formidability did not consistently predict subjective ratings of coalitional formidability. This is unexpected because less formidable individuals overestimated objective coalition size in other tasks, and coalition size as such predicted formidability ratings. One interpretation is that automatic processes, such as those engaged under brief exposure, prompt caution-driven overestimations, while explicit judgments about strength rely on different cues and assessment processes. However, these interpretations should be treated with caution. The sample of Study 7 was relatively small. Our power analyses focused on trial-level variables, utilizing repeated observations within participants, but predictors like individual formidability vary only across participants. Their effective sample size is thus limited to the number of participants, not observations (as noted in SI5, power calculations targeted main effects of target categories, not individual-level predictors).

Perceived ingroup coalitional support did not predict higher threat ratings or overestimation errors. This may also be attributed to the limited sample (this variable did not vary within individuals), but also to measurement: the scale we used assessed general rather than situational coalitional support. According to our theoretical model, situational variables (e.g., immediate presence of allies during encounter) would be more suitable to test the effects on coalitional assessments. Due to the constraints of online administration, we relied on trait-like proxies. More direct manipulations, including ally presence/absence during the task, were implemented in separate studies in Ukraine (reported in separate publication).

Effect sizes across Studies 1–7 were modest. We intentionally adopted a conservative approach, using highly schematic stimuli and omitting contextual cues. The adaptations we hypothesize would be expected to operate most strongly in naturalistic, high-stakes intergroup encounters. If they nonetheless produce measurable effects under such stripped-down conditions, they are likely even more potent in real-world intergroup encounters. Furthermore, the effect sizes observed here are consistent with those reported in prior work using the same paradigms. Existing studies report 10ms difference as substantively meaningful in the dot-probe analyses (Chapman *et al.*, 2019; Durkee *et al.*, 2018; Staugard, 2009) and 15ms is used as a reference in power analyses in reaction-time experiments more generally (Brysbaert & Stevens, 2018).

Considering alternative explanations, the attentional bias toward coordinated male-like groups cannot be attributed to humanness alone, as all stimuli featured human-like targets. Nor can it be reduced to male features per se: rotated male-like targets did not elicit the same attentional bias. Coordination alone also fails to explain the effect: coordinated female-like groups did not produce similar biases. The enumeration findings similarly surpassed what would be expected from the similarity grouping effect: coordinated male-like groups were enumerated more accurately than both rotated male-like targets and coordinated female-like groups. The counting rapidity experiments provided less definitive evidence: the enumeration advantage only persisted under higher task demands (i.e., with target N = 8). These latter tests were not preregistered, but their results are consistent with multiple pilot studies detailed in SI, showing considerably larger effect sizes, across multiple tasks, for stimuli with 8 targets.

We also considered socio-cultural explanations. One possibility is that people learn to attend to threatening coalitions through developmental exposure to group-based violence. If this were true, individuals who perceive greater intergroup hostility in their environment should show stronger responses in our tasks (e.g., faster attention to coordinated male-like groups). To explore this, we included a 3-item scale measuring generalized group threat perceptions (e.g., "I feel safe walking past a group of strangers"). However, this measure showed no clear relationship with task performance. A more comprehensive evaluation, such as cross-cultural comparison of higher-level units (e.g., countries) with varying histories of intergroup conflict, is needed to more fully assess socio-cultural alternatives (see below).

The evidence supporting 11 of the 14 hypotheses should be considered in combination. While alternative explanations might account for individual findings, such as why participants attended to coordinated male-like groups vs non-coordinated ones, it is far harder to propose a single alternative that explains the full pattern: attentional biases, accurate and rapid enumeration, and systematic variation in subjective formidability ratings. Our interpretation treats these results as interconnected outputs of the same underlying adaptation. It is even harder to develop such an integrated explanation a priori rather than post hoc, after seeing the data. By contrast, our hypotheses were derived a priori through a systematic adaptationist framework and preregistered before data collection.

Together, these findings—automatic attention to coordinated male-like coalitions, more accurate and rapid enumeration by male participants, threat ratings that scale with coalition size and individual size, and systematic overestimation by females and less formidable participants—point to a coherent pattern. This pattern supports the existence of specialized evolved mechanisms—i.e., *special design* (Williams, 1966)—for detecting, evaluating, and responding to coalitional threats. Viewed as an integrated whole, the evidence aligns with the broader hypothesis that small-scale war was a recurrent feature of human evolutionary history.

The final step in evaluating the overarching hypothesis of adaptation to war lies in cross-cultural testing. We are conducting Study 8 that replicates the core experimental paradigms—automatic attention, enumeration accuracy, and coalitional formidability assessments—across 30 countries using representative samples in a multi-level framework. By testing

whether the observed patterns hold across diverse cultural contexts, this research will allow us to assess the universality of these mechanisms. If the effects observed in Studies 1–7 do not significantly vary across countries, it would provide strong evidence that these tendencies reflect evolved, species-typical adaptations shaped by ancestral intergroup aggression rather than culturally contingent learning.

## Ethics and consent

The research consists of Studies 1–7, all of which were administered in the United States and had been review exempted due to minimal risk by the US-based WCG IRB. Informed consent was obtained from all participants digitally. The participants were compensated in US dollars following the recommendations of the online recruitment platform Prolific we utilize in our research. We paid the participants the minimal hourly rate as required by Prolific, which was 8 USD, plus 50% of this rate, totaling 12 USD/hour. The actual payments varied across Studies 1–7, because completion times were different (ranging from 4 to 12 minutes).

## Data availability

The data collected as part of Studies 1–7 are publicly available online on the Open Science Framework (OSF) repository associated with the project: https://doi.org/10.17605/OSF.IO/RYV4N (Bartusevičius *et al.*, 2025a).

The data collected as part of Pilots 1, 2.1–2.8, 3.1–3.16 (Bartusevičius *et al.*, 2025b) are also publicly available online on the same repository: https://doi.org/10.17605/OSF.IO/RYV4N.

This project contains the following data:

- AWAR_WP2_SI 4.1 (Protocol S1-7)_preregistration_version_final.pdf
- Pilots 1–3 of Work Package 2 of AWAR project
- Supporting Information for Bartusevičius *et al.*, 2026 ORE
- Studies 1–7 of Work Package 2 of AWAR project

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

## Software availability

The analysis code for Studies 1–7 is publicly available online on the Open Science Framework (OSF) repository associated with the project.

- Source code available from: https://osf.io/ryv4n/files/osfstorage
- Archived software available from: 10.17605/OSF.IO/RYV4N
- License: CC-By Attribution 4.0 International

### Acknowledgements

Earlier stages of the project were presented at 2021, 2023, and 2024 Annual Conferences of the Human Behavior and Evolution Society, 2023 International Congress of International Society for Human Ethology, 2023 Annual Meeting of Society for Neuroscience, 2023 Annual Meeting of Psychonomic Society, internal seminars at the Center for Evolutionary Psychology (University of California, Santa Barbara), the Center for the Study of Human Nature (California State University, Fullerton), the Experimental Biological Anthropology (XBA) Lab (University of California, Los Angeles), the Culture, Society, and Behavior (CSB) Lab (University of Oslo), the Evolutionary Social Psychology (KNVB) Co-Laboratory (Arizona State University), and the Peace Research Institute Oslo. We thank the participants for their feedback.

## References

Bartusevičius H: **Physical strength predicts political violence.** *Evol Hum Behav.* 2021; **42**(5): 423–430.
**Publisher Full Text**

Bartusevičius H, Aminihajibashi S, Goetz S, *et al.*: **Data for: Studies 1-7 of work package 2 of AWAR project.** *Open Science Framework*. 2025a.
**http://www.doi.org/10.17605/OSF.IO/RYV4N**

Bartusevičius H, Aminihajibashi S, Goetz S, *et al.*: **Data for: Pilots 1-3 of work package 2 of AWAR project.** *Open Science Framework*. 2025b.
**http://www.doi.org/10.17605/OSF.IO/RYV4N**

Boehm C: **Ancestral hierarchy and conflict.** *Science.* 2012; **336**(6083): 844–847.
**PubMed Abstract** | **Publisher Full Text**

Boehm C: **The biocultural evolution of conflict resolution between groups**. In: Fry DP (ed) *War, peace, and human nature*. Oxford University Press, 2015.
**Publisher Full Text**

Bowles S: **Group competition, reproductive leveling, and the evolution of human altruism.** *Science.* 2006; **314**(5805): 1569–1572.
**PubMed Abstract** | **Publisher Full Text**

Bowles S: **Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors?** *Science.* 2009; **324**(5932): 1293–1298.
**PubMed Abstract** | **Publisher Full Text**

Brysbaert M, Stevens M: **Power analysis and effect size in mixed effects models: a tutorial.** *J Cogn.* 2018; **1**(1): 9.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Chapman A, Devue C, Grimshaw GM: **Fleeting reliability in the dot-probe task.** *Psychol Res.* 2019; **83**(2): 308–320.
**PubMed Abstract** | **Publisher Full Text**

Choi JK, Bowles S: **The coevolution of parochial altruism and war.** *Science.* 2007; **318**(5850): 636–640.
**PubMed Abstract** | **Publisher Full Text**

Chomsky N: **Excerpt from book cover.** In: Fry DP (ed) *War, peace, and human nature*. Oxford University Press, 2015.

Cisler JM, Koster EHW: **Mechanisms of attentional biases towards threat in anxiety disorders: an integrative review.** *Clin Psychol Rev.* 2010; **30**(2): 203–216.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Cosmides L, Tooby J, Barkow JH: **Introduction: evolutionary psychology and**

**conceptual integration**. In: Barkow JH, Cosmides L, & Tooby J (eds) *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press, 1992.
**Publisher Full Text**

Cosmides L, Tooby J: **The theoretical foundations of evolutionary psychology**. In: Buss DM (ed) *The handbook of evolutionary psychology*. Second edition. John Wiley & Sons, 2015.
**Publisher Full Text**

Crump MJC, McDonnell JV, Gureckis TM: **Evaluating Amazon's mechanical turk as a tool for experimental behavioral research.** *PLoS One.* 2013; **8**(3): e57410.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Diab K: **Natural born killers**. 2008.
**Reference Source**

Durkee PK, Goetz AT, Lukaszewski AW: **Formidability assessment mechanisms: examining their speed and automaticity.** *Evol Hum Behav.* 2018; **39**(2): 170–178.
**Publisher Full Text**

Ember CR: **Myths about hunter-gatherers.** *Ethnology.* 1978; **17**(4): 439–448.
**Publisher Full Text**

Evans TC, Britton JC: **Improving the psychometric properties of dot-probe attention measures using response-based computation.** *J Behav Ther Exp Psychiatry.* 2018; **60**: 95–103.
**PubMed Abstract** | **Publisher Full Text**

Ferguson RB: **Born to live: challenging killer myths**. In: Sussman R & Cloninger C (eds) *Origins of altruism and cooperation*. *Developments in primatology: Progress and prospects*. Springer, 2011.
**Publisher Full Text**

Ferguson RB: **The prehistory of war and peace in Europe and the Near East**. In: Fry DP (ed) *War, peace, and human nature*. Oxford University Press, 2015.
**Publisher Full Text**

Fessler DMT, Holbrook C: **Friends shrink foes: the presence of comrades decreases the envisioned physical formidability of an opponent.** *Psychol Sci.* 2013a; **24**(5): 797–802.
**PubMed Abstract** | **Publisher Full Text**

Fessler DMT, Holbrook C: **Bound to lose: physical incapacitation increases the conceptualized size of an antagonist in men.** *PLoS One.* 2013b; **8**(8): e71306.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Fessler DMT, Holbrook C: **Marching into battle: synchronized walking diminishes the conceptualized formidability of an antagonist in men.** *Biol Lett.* 2014; **10**(8): 20140592.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Fessler DMT, Holbrook C: **Synchronized behavior increases assessments of the formidability and cohesion of coalitions.** *Evol Hum Behav.* 2016; **37**(6): 502–509.
**Publisher Full Text**

Fessler DMT, Holbrook C, Dashoff D: **Dressed to kill? Visible markers of coalitional affiliation enhance conceptualized formidability.** *Aggress Behav.* 2016; **42**(3): 299–309.
**PubMed Abstract** | **Publisher Full Text**

Fessler DMT, Holbrook C, Gervais MM: **Men's physical strength moderates conceptualizations of prospective foes in two disparate societies.** *Hum Nat.* 2014; **25**(3): 393–409.
**PubMed Abstract** | **Publisher Full Text**

Fessler DMT, Holbrook C, Snyder JK: **Weapons make the man (larger): formidability is represented as size and strength in humans.** *PLoS One.* 2012; **7**(4): e32751.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Fessler DMT, Holbrook C, Pollack JS, *et al.*: **Stranger danger: parenthood increases the envisioned bodily formidability of menacing men.** *Evol Hum Behav.* 2014a; **35**(2): 109–117.
**Publisher Full Text**

Fessler DMT, Tiokhin LB, Holbrook C, *et al.*: **Foundations of the crazy bastard hypothesis: nonviolent physical risk-taking enhances conceptualized formidability.** *Evol Hum Behav.* 2014b; **35**(1): 26–33.
**Publisher Full Text**

Fink B, Neave N, Seydel H: **Male facial appearance signals physical strength to women.** *Am J Hum Biol.* 2007; **19**(1): 82–87.
**PubMed Abstract** | **Publisher Full Text**

Fisher RS, Acharya JN, Baumer FM, *et al.*: **Visually sensitive seizures: an updated review by the epilepsy foundation.** *Epilepsia.* 2022; **63**(4): 739–768.
**PubMed Abstract** | **Publisher Full Text**

Fry DP: **The human potential for peace: an anthropological challenge to assumptions about war and violence**. Oxford University Press, 2006.
**Reference Source**

Fry DP: **Beyond war: the human potential for peace**. Oxford University Press, 2007.
**Publisher Full Text**

Fry DP: **War, peace, and human nature**. Oxford University Press, 2015.
**Reference Source**

Fry DP, Söderberg P: **Lethal aggression in mobile forager bands and implications for the origins of war.** *Science.* 2013; **341**(6143): 270–273.
**PubMed Abstract** | **Publisher Full Text**

Fuentes A: **The creative spark: how imagination made humans exceptional**. Dutton, 2017.
**Reference Source**

Fukuyama F: **Women and the evolution of world politics.** *Foreign Aff.* 1998; **77**(5): 24–40.
**Publisher Full Text**

Gat A: **War in human civilization**. Oxford University Press, 2008.
**Reference Source**

Glowacki L: **How the tribal warfare of our ancestors explains the Islamic State**. 2016.
**Reference Source**

Goodall J: **The chimpanzees of gombe: patterns of behavior**. Harvard University Press, 1986.
**Reference Source**

Gorman J: **Prehistoric massacre hint at war among hunter-gatherers**. 2016.
**Reference Source**

Gomez JM, Verdú M, González-Megías A, *et al.*: **The phylogenetic roots of human lethal violence.** *Nature.* 2016; **538**(7624): 233–237.
**PubMed Abstract** | **Publisher Full Text**

Goldstein JS: **Winning the war on war.** Dutton, 2011.
**Reference Source**

Guilaine J, Zammit J: **The origins of war: violence in prehistory**. Blackwell Publishing, 2005.
**Publisher Full Text**

Haas J, Piscitelli M: **The prehistory of warfare: misled by ethnography**. In: Fry DP (ed) *War, peace, and human nature.* Oxford University Press, 2015.
**Reference Source**

Holbrook C, Fessler DMT, Navarrete CD: **Looming large in others' eyes: racial stereotypes illuminate dual adaptations for representing threat versus prestige as physical size.** *Evol Hum Behav.* 2016a; **37**(1): 67–78.
**Publisher Full Text**

Holbrook C, Fessler DMT, Pollack J: **With God on our side: religious primes reduce the envisioned physical formidability of a menacing adversary.** *Cognition.* 2016b; **146**: 387–392.
**PubMed Abstract** | **Publisher Full Text**

Holbrook C, López-Rodríguez L, Fessler DMT, *et al.*: **Gulliver's politics: conservatives envision potential enemies as readily vanquished and physically small.** *Soc Psychol Pers Sci.* 2017; **8**(6): 670–678.
**Publisher Full Text**

Horgan J: **War scholar critiques new study of roots of violence**. 2016.
**Reference Source**

Johnson CY: **Study suggests violence is an evolutionary adaptation**. 2014.
**Reference Source**

Keeley LH: **War before civilization**. Oxford University Press, 1996.
**Reference Source**

Kelly RC: **Warless societies and the origin of war.** University of Michigan Press, 2000.
**Reference Source**

Kelly RC: **The evolution of lethal intergroup violence.** *Proc Natl Acad Sci U S A.* 2005; **102**(43): 15294–15298.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Kochari AR: **Conducting web-based experiments for numerical cognition research.** *J Cogn.* 2019; **2**(1): 39.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Lahr MM, Rivera F, Power RK, *et al.*: **Inter-group violence among early Holocene hunter-gatherers of West Turkana, Kenya.** *Nature.* 2016; **529**(7586): 394–398.
**PubMed Abstract** | **Publisher Full Text**

LeBlanc SA: **Prehistoric warfare in the American Southwest.** University of Utah Press, 1999.
**Reference Source**

LeBlanc SA, Register KE: **Constant battles: the myth of the peaceful, noble savage.** St. Martin's Press, 2003.
**Reference Source**

Lewis DMG, Al-Shawaf L, Conroy-Beam D, *et al.*: **Evolutionary psychology: a how-to-guide.** *Am Psychol.* 2017; **72**(4): 353–373.
**PubMed Abstract** | **Publisher Full Text**

Lieberman D, Tooby J, Cosmides L: **The architecture of human kin detection.** *Nature.* 2007; **445**(7129): 727–731.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Loomis AM, Evans TC, Grasso DJ, *et al.*: **Improved reliability of dot probe measures with response-based computation: an application with young violence-exposed children.** *Cogn Ther Res.* 2023; **47**: 968–979.
**Publisher Full Text**

Macleod C, Mathews A, Tata P: **Attentional bias in emotional disorders.** *J Abnorm Psychol.* 1986; **95**(1): 15–20.
**PubMed Abstract** | **Publisher Full Text**

Manson JH, Wrangham RW, Boone JL, *et al.*: **Intergroup aggression in chimpanzees and humans.** *Curr Anthropol.* 1991; **32**(4): 369–390.
**Publisher Full Text**

Marlowe FW: **Hunter-gatherers and human evolution.** *Evol Anthropol.* 2005; **14**(2): 54–67.
**Publisher Full Text**

Marr D: **Vision: a computational investigation into the human representation and processing of visual information.** WH Freeman and Company, 1982.
**Reference Source**

Martin DL, Harrod RP: **Bioarcheological contributions to the study of violence.** *Am J Phys Anthropol.* 2015; **156 Suppl 59**: 116–145.
**PubMed Abstract** | **Publisher Full Text**

Mayr E: **How to carry out the adaptationist program?** *Am Nat.* 1983; **121**(3): 324–334.
**Publisher Full Text**

McCormick PA: **Orienting attention without awareness.** *J Exp Psychol Hum Percept Perform.* 1997; **23**(1): 168–180.
**PubMed Abstract** | **Publisher Full Text**

Mogg K, Millar N, Bradley BP: **Biases in eye movements to threatening facial expressions in generalized anxiety disorder and depressive disorder.** *J Abnorm Psychol.* 2000; **109**(4): 695–704.
**PubMed Abstract** | **Publisher Full Text**

Obama BH: **Nobel lecture**. 2009.
**Reference Source**

Otterbein KF: **The evolution of war: a cross-cultural study.** HRAF Press, 1970.
**Reference Source**

Otterbein KF: **How war began.** Texas A&M University Press, 2004.
**Reference Source**

Peer E, Brandimarte L, Samat S, *et al.*: **Beyond the Turk: alternative platforms for crowdsourcing behavioral research.** *J Exp Soc Psychol.* 2017; **70**: 153–163.
**Publisher Full Text**

Peer E, Rothschild D, Gordon A, *et al.*: **Data quality of platforms and panels for online behavioral research.** *Behav Res Methods.* 2022; **54**(4): 1643–1662.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Pinker S: **The better angels of our nature**. Penguin, 2011a.
**Reference Source**

Pinker S: **Violence vanquished**. 2011b.

Price ME, Dunn J, Hopkins S, *et al.*: **Anthropometric correlates of human anger.** *Evol Hum Behav.* 2012; **33**(3): 174–181.
**Publisher Full Text**

Romm C: **A prehistoric mass grave suggests hunter-gatherers weren't so peaceful**. 2016.
**Reference Source**

Sauter M, Draschkow D, Mack W: **Building, hosting and recruiting: a brief introduction to running behavioral experiments online.** *Brain Sci.* 2020; **10**(4): 251.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Schleifer P, Landerl K: **Subitizing and counting in typical and atypical development.** *Dev Sci.* 2011; **14**(2): 280–291.
**PubMed Abstract** | **Publisher Full Text**

Sell A, Bryant GA, Cosmides L, *et al.*: **Adaptations in humans for assessing physical strength from the voice.** *Proc Biol Sci.* 2010; **277**(1699): 3509–3518.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Sell A, Sznycer D, Cosmides L, *et al.*: **Physically strong men are more militant: a test across four countries.** *Evol Hum Behav.* 2017; **38**(3): 334–340.
**Publisher Full Text**

Sell A, Tooby J, Cosmides L: **Formidability and the logic of human anger.** *Proc Natl Acad Sci U S A.* 2009; **106**(35): 15073–15078.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Shostak M: **Nisa, the life and words of a !Kung woman**. Harvard University Press, 1981.
**Reference Source**

Staugard SR: **Reliability of two versions of the dot-probe task using photographic faces.** *Psychol Sci.* 2009; **51**(3): 339–350.
**Reference Source**

Stojanowski CM, Seidel AC, Fulginiti LC, *et al.*: **Contesting the massacre at Nataruk.** *Nature.* 2016; **539**(7630): E8–E10.
**PubMed Abstract** | **Publisher Full Text**

Suwa G, Kono RT, Simpson SW, *et al.*: **Paleobiological implications of the *Ardipithecus ramidus* dentition.** *Science.* 2009; **326**(5949): 94–99.
**PubMed Abstract** | **Publisher Full Text**

The New York Times: **Text of President Obama's speech in Hiroshima, Japan**. 2016.
**Reference Source**

Thigpen NN, Gruss LF, Garcia S, *et al.*: **What does the dot-probe task measure? A reverse correlation analysis of electrocortical activity.** *Psychophysiology.* 2018; **55**(6): e13058.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Tooby J, Cosmides L: **The evolution of war and its cognitive foundations.** *Evol Hum Behav.* 2025; **46**(3): 106687.
**Publisher Full Text**

Trick LM, Pylyshyn ZW: **Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision.** *Psychol Rev.* 1994; **101**(1): 80–102.
**PubMed Abstract** | **Publisher Full Text**

Vencl S: **War and warfare in archaeology.** *J Anthropol Archaeol.* 1984; **3**(2): 116–132.
**Publisher Full Text**

von Rueden C, Gurven M, Kaplan H: **The multiple dimensions of male social status in an Amazonian society.** *Evol Hum Behav.* 2008; **29**(6): 402–415.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Walker PL: **A Bioarchaeological perspective on the history of violence.** *Annual Review of Anthropology.* 2001; **30**: 573–596.
**Publisher Full Text**

White RD, Asfaw B, Beyene Y, *et al.*: ***Ardipithecus ramidus* and the paleobiology of early hominids.** *Science.* 2009; **326**(5949): 75–86.
**PubMed Abstract** | **Publisher Full Text**

Williams GC: **Adaptation and natural selection**. Princeton University Press, 1966.

Wilson ML, Boesch C, Fruth B, *et al.*: **Lethal aggression in *Pan* is better explained by adaptive strategies than human impacts.** *Nature.* 2014; **513**(7518): 414–417.
**PubMed Abstract** | **Publisher Full Text**

Wilson ML, Wrangham RW: **Intergroup relations in chimpanzees.** *Annual Review of Anthropology.* 2003; **32**(1): 363–392.
**Publisher Full Text**

Wobst HM: **The archaeo-ethnology of hunter-gatherers or the tyranny of the ethnographic record in archeology.** *American Antiquity.* 1978; **43**(2): 303–309.
**Publisher Full Text**

Wrangham RW: **Evolution of coalitionary killing.** *Am J Phys Anthropol.* 1999; **110**(Suppl 29): 1–30.
**PubMed Abstract** | **Publisher Full Text**

Wrangham RW, Glowacki L: **Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers: evaluating the chimpanzee model.** *Hum Nat.* 2012; **23**(1): 5–29.
**PubMed Abstract** | **Publisher Full Text**

Wrangham RW, Peterson D: **Demonic males: Apes and the origins of human violence**. Houghton Mifflin, 1996.
**Reference Source**

Yu D, Xiao X, Bemis DK, *et al.*: **Similarity grouping as feature-based selection.** *Psychol Sci.* 2019; **30**(3): 376–385.
**PubMed Abstract** | **Publisher Full Text**

# Open Peer Review

## Current Peer Review Status: ✗ ✓ ?

**Version 1**

Reviewer Report 14 August 2026

https://doi.org/10.21956/openreseurope.23730.r72117

© **2026 Schumacher G.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

**Gijs Schumacher**
University of Amsterdam, Amsterdam, The Netherlands

This is a very interesting but also a paper very dense in information, especially for someone's whose core expertise lies outside evolutionary psychology. Let me tell you how I read it and deduce some recommendations from it.

The literature review section is really interesting. I cannot judge how accurately the literature is reflected in this review. So, I am curious to see how the authors respond to the comments of reviewer 1 regarding this. At the same time, I also feel that the literature review is somewhat disconnected from the article's research question. Whether war was a recurrent feature in the Pleistocene, or not, is perhaps not the most relevant issue. For one because conflict isn't exclusively intergroup conflict (it could be interpersonal, intragroup or with animals), or perhaps alternatively even very rare experiences with war may already be sufficient for adaptations to develop, because of its potentially disastrous effects on group well-being.

After the literature review there is a quick turn towards declaring interest for coalitional violence. But from there it seems to be exclusively intergroup conflict that drives the development of mechanisms that cope with war, and the possibility of intragroup conflict (equally disastrous) and animal attacks (also acting in coalition) is not discussed.
The explanation of the tasks and hypotheses is quite clear to me, although the language can be quite technical and dense.

The materials and methods section were really quite hard to follow. It is very, very long. I think all the info on the pilot studies can go to the appendix, also because these findings are not very surprising. I wasn't sure about the preliminary tests H1-H14 section. I didn't understand how uncertainty was quantified and translated into categorical statements. Also, if the authors think these tests may be underpowered then what is their purpose, especially if the exploratory tests weren't used for hypothesis selection.
In the analysis plan some pieces of text are almost identical, with only the dependent variable description changing. This can be written down more efficiently (e.g predictions 7-9 have identical texts except for the dv).

The results section was also hard to follow. Minimally I would exclude all the non-preregistered checks here. Would it be possible to say a bit more about how large these effects are?

Table 1 is great!

Finally, the discussion does not seem to relate to the wider literature. Almost none of the literature cited in the frontend returns here.

**Is the work clearly and accurately presented and does it engage with the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**Are all the source data and materials underlying the results available?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Yes

***Competing Interests:*** No competing interests were disclosed.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 14 April 2026

https://doi.org/10.21956/openreseurope.23730.r71420

**© 2026 Sell A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Aaron Sell**
Heidelberg University, Tiffin, Ohio, USA

The paper "Evidence from over 30 studies suggests the existence of evolved adaptations to war in humans" providences' evidence in the ongoing debate about the evolutionary history of war.

Debates about pre-historical behaviors are difficult because of the absence of conclusive data from archeology, and so the authors use evolutionary theory to predict mental adaptations. In short, if human ancestors engaged in war repeatedly, we should have evolved mechanisms that function in that environment. This is - to my mind - extremely sound reasoning.

The authors then go on to describe over twenty exploratory studies and 7 key pre-registered studies that show evidence that the human mind has been designed by evolution to respond to coalitional threats from men (something that would have likely evolved if we came from an evolutionary history of coalitional aggression).

I think the paper was extremely methodical, the literature review clear and comprehensive given space constraints. The predictions were reasonable and good tests of the theory, and the results persuasive.

Overall, it is an excellent paper and I recommend acceptance.

I have a few minor comments that I will put only as suggestions for the authors.

1. Discussion of the numerous pilots was difficult to follow given how many there were. I'm wondering if the manuscript wouldn't be cleaner if the pilot data was made less detailed and merely summarized.

2. Given the number of studies, it was hard to keep all seven methods in mind as the paper transitions from Methods to Results. I wonder if the paper wouldn't be easier to read if one study was reported at a time, including the methods and its results, before moving on to the next Study.

**Is the work clearly and accurately presented and does it engage with the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**Are all the source data and materials underlying the results available?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
I cannot comment. A qualified statistician is required.

**Are the conclusions drawn adequately supported by the results?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Evolutionary psychology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 11 April 2026

https://doi.org/10.21956/openreseurope.23730.r71416

© **2026 Lopez A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Anthony Lopez**
Washington State University, Pullman, Washington, USA

This research tests a proposition: “If war recurred during evolution, then our minds are likely equipped with efficient, complex, and specialized mechanisms.” The authors propose that if such a suite of systems exist, one necessary component of these systems would be “the existence of rapid and accurate coalition detection, enumeration, and formidability assessment mechanisms.” They then perform a series of experiments to determine if such mechanisms exist, which they take as evidence of adaptation for war. My review is simple: while the starting proposition is sound, and detection/assessment mechanisms should indeed be a component of resultant war adaptations, I disagree that evidence of such mechanisms is, on its own, strong or unique evidence of adaptations for war. The strongest claim I think the authors can make based on their findings is that they have suggestive evidence in support of design for coalitional competition, and secondarily, although a stretch, evidence of raid adaptations. I elaborate below.

First, the literature review is too expansive for what the authors ultimately test. Furthermore, its structure and exposition, closely mirrors already well-known and effective surveys of this question (Kissel and Kim 2018; Gat 2015; Lopez 2020; Hames 2019; Glowacki et al. 2017). It’s not clear why this all needs to be rehashed. The intro can be significantly condensed.

Second, the authors' data do not speak to adaptations for war; rather, they test only a much narrower proposition: Given the real likelihood of violent ambush in ancestral contexts, what adaptations ought to exist to facilitate survival from them? The empirical design maps onto a narrower adaptive problem—ambush detection and response—rather than the broader construct of war. I would encourage the authors to limit the scope of their claims to align with the adaptive problem their empirical design actually tests. It is a fair leap from adaptations for ambush survival to adaptations for war.

Third, the authors, via significant omission, seem to suggest we do not yet have evidence of adaptations for coalitional violence. The opposite is true. For starters, see: several studies by Mark van Vugt and colleagues on the Male Warrior Hypothesis (which by the way relate directly to the authors’ own hypotheses); several studies by Tooby, Cosmides, Kurzban, as well as Pietraszewski, and also work by Michael Bang Petersen, relating to coalitional detection (again, relating directly to the authors’ own hypotheses); Glowacki and McDermott on key individuals in initiating raids; a voluminous literature on stable offense/defense psychology by scholars such as De Dreu, Lopez, Rusch; several studies on leadership in coalitional conflict by Spisak, Little, Petersen and many

others; work on overconfidence in the context of intergroup violence by Dominic Johnson and Wrangham. If none of these were relevant to the specific hypotheses the authors want to investigate, the omission would be less serious, but in fact many of them directly relate.

Fourth, the authors claim to have performed a task analysis, but it feels thin in places. The adaptive problem itself (surviving ambush) seems to collapse at least two very different dimensions (avoidance and deterrence), and some of the hypotheses seem to jump from broad trends into hypotheses with little elaboration (e.g. sex differences). The proposed task analysis does not fully specify the computational structure of the problem (inputs, decision rules, error asymmetries).

Fifth, I offer a series of specific page-related comments below:

“The debate essentially revolves around the question of whether war was a recurrent feature of life in the Pleistocene” (p. 3). It does not hinge on questions of recurrence alone, but also reproductive significance. This is not a minor point.

“then one of the most consequential problems ancestral hominins faced was how not to fall victim” (p. 7) Awkward phrasing. “How not to fall victim” has at least two dimensions: avoidance and deterrence. Either the authors collapse the two, or they are choosing one without saying it. These are very different problems with likely very different solutions. In fact, this would be a profitable distinction for the authors to investigate, particularly since there has been much written recently on evolution of revenge and deterrence, which may relate to raid dynamics (e.g. McCullough 2013; McDermott et al. 2017).

“are beyond our research focus” (p. 8) I find this conclusion both surprising and interesting. Could the authors say why?

“People enumerate, accurately (H2) and rapidly (H3), members of potentially threatening coalitions” (p. 9) Why only threatening coalitions? If we are truly adapted for raiding, we should be scanning opportunistically not selectively or defensively.

“Males, compared to females, determine the size of potentially threatening coalitions more accurately (H9) and rapidly (H10) When females make errors in these estimates, they tend to overestimate more so than males (H11) Females, compared to males, subjectively assess potentially threatening coalitions as more formidable (H12)” (p. 10) The authors should spell out in greater detail how we go from coalitional violence as “a male activity” to landing at these four hypotheses. Again, some of the omitted literature I reference above could have been helpful.

“we examined how cues of ingroup/outgroup membership assigned to targets influence enumeration performance (P3.13–P3.16), but this manipulation was discontinued for the reasons explained on p. 37 in SI3” (p. 17) This would have been an interesting angle to investigate.

“1 was supported with low uncertainty” (p. 18) I don’t think a verbal assertion of patterns suffices for a presentation of results. Especially since, in my opinion, one pattern that really stands out across findings is the generation of associations that are significant, but substantively fairly small. Perhaps some discussion and/or display of effect sizes? I don’t really know what qualifies as “supportive” or low vs high uncertainty in the authors eyes. For example, would “supported with

high uncertainty" just be "not supported"? I'm very confused by this presentation.

"Effect sizes across Studies 1–7 were modest. We intentionally adopted a conservative approach, using highly schematic stimuli and omitting contextual cues. The adaptations we hypothesize would be expected to operate most strongly in naturalistic, high-stakes intergroup encounters. If they nonetheless produce measurable effects under such stripped-down conditions, they are likely even more potent in real-world intergroup encounters" (p. 43) The authors are attempting to make an *a fortiori* argument, but I think this needs to be demonstrated, not assumed. The authors seem to claim a form of mismatch hypothesis in which adaptations produce "stronger" signals in EEA or ecologically valid environments, but "weaker" signals in novel, modern or 'stripped-down conditions.' I don't see why that should necessarily be the case, but perhaps I am misunderstanding. I would appreciate being walked through the logic.

**References**

1. Gat A: Proving communal warfare among hunter-gatherers: The quasi-rousseauan error. *Evolutionary Anthropology: Issues, News, and Reviews*. 2015; **24** (3): 111-126 Publisher Full Text
2. Glowacki L, Wilson M, Wrangham R: The evolutionary anthropology of war. *Journal of Economic Behavior & Organization*. 2020; **178**: 963-982 Publisher Full Text
3. Hames R: Pacifying Hunter-Gatherers. *Human Nature*. 2019; **30** (2): 155-175 Publisher Full Text
4. McDonald M, Navarrete C, Van Vugt M: Evolution and the psychology of intergroup conflict: the male warrior hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2012; **367** (1589): 670-679 Publisher Full Text
5. Vugt M, Cremer D, Janssen D: Gender Differences in Cooperation and Competition. *Psychological Science*. 2007; **18** (1): 19-23 Publisher Full Text
6. Kurzban R, Tooby J, Cosmides L: Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*. 2001; **98** (26): 15387-15392 Publisher Full Text
7. Cosmides L, Tooby J, Kurzban R: Perceptions of race. *Trends in Cognitive Sciences*. 2003; **7** (4): 173-179 Publisher Full Text
8. Pietraszewski D: Toward a computational theory of social groups: A finite set of cognitive primitives for representing any and all social groups in the context of conflict. *Behavioral and Brain Sciences*. 2022; **45**. Publisher Full Text
9. Glowacki L, McDermott R: Key individuals catalyse intergroup violence. *Philosophical Transactions of the Royal Society B*. 2022; **377** (1851). Publisher Full Text
10. De Dreu C, Gross J: Revisiting the form and function of conflict: Neurobiological, psychological, and cultural mechanisms for attack and defense within and between groups. *Behavioral and Brain Sciences*. 2019; **42**. Publisher Full Text
11. De Dreu C, Gross J, Méder Z, Giffin M, et al.: In-group defense, out-group aggression, and coordination failures in intergroup conflict. *Proceedings of the National Academy of Sciences*. 2016; **113** (38): 10524-10529 Publisher Full Text
12. Rusch H: The evolutionary interplay of intergroup conflict and altruism in humans: a review of parochial altruism theory and prospects for its extension. *Proceedings of the Royal Society B: Biological Sciences*. 2014; **281** (1794). Publisher Full Text
13. Spisak B, O'Brien M, Nicholson N, van Vugt M: Niche Construction and the Evolution of Leadership. *Academy of Management Review*. 2015; **40** (2): 291-306 Publisher Full Text
14. Castro P, Seixas E, Neca P, Bettencourt L: Successfully Contesting the Policy Sphere: Examining Through the Press a Case of Local Protests Changing New Ecological Laws. *Political Psychology*. 2018; **39** (1): 107-123 Publisher Full Text

15. Johnson D, McDermott R, Barrett E, Cowden J, et al.: Overconfidence in wargames: experimental evidence on expectations, aggression, gender and testosterone. *Proceedings of the Royal Society B: Biological Sciences*. 2006; **273** (1600): 2513-2520 Publisher Full Text
16. Johnson D, Fowler J: The evolution of overconfidence. *Nature*. 2011; **477** (7364): 317-320 Publisher Full Text
17. Johnson D, Weidmann N, Cederman L: Fortune Favours the Bold: An Agent-Based Model Reveals Adaptive Advantages of Overconfidence in War. *PLoS ONE*. 2011; **6** (6). Publisher Full Text
18. Kim N, Kissel M: Emergent Warfare in our Evolutionary Past. 2018. Publisher Full Text
19. Anthony C. Lopez: The Evolution of Warfare. *https://www.taylorfrancis.com/chapters/edit/10.4324/9780429242801-2/the-evolution-of-warfare-anthony-c-lopez*. 2020.
20. Spisak B, Homan A, Grabo A, Van Vugt M: Facing the situation: Testing a biosocial contingency model of leadership in intergroup relations using masculine and feminine faces. *The Leadership Quarterly*. 2012; **23** (2): 273-280 Publisher Full Text
21. McCullough M, Kurzban R, Tabak B: Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences*. 2013; **36** (1): 1-15 Publisher Full Text
22. R, McDermott A, Lopez P, Hatemi: 'Blunt Not the Heart, Enrage It': The Psychology of Revenge and Deterrence. *https://tnsr.org/2017/11/blunt-not-heart-enrage-psychology-revenge-deterrence/*. 2017.
23. Lopez A: The Evolutionary Psychology of War: Offense and Defense in the Adapted Mind. *Evolutionary Psychology*. 2017; **15** (4). Publisher Full Text

**Is the work clearly and accurately presented and does it engage with the current literature?**
No

**Is the study design appropriate and is the work technically sound?**
No

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**Are all the source data and materials underlying the results available?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Partly

**Are the conclusions drawn adequately supported by the results?**
No

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** evolutionary psychology, international relations, psychology of war, collective action, leadership, deterrence, offense/defense dynamics.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for**

**reasons outlined above.**