Unveiling the associative mechanisms underlying the additive bias: using an Implicit Association Test to gain insight into people's preference for additive actions.

Maria Adriana Neroni[1], Nathan Crilly[2] and Maria Antonella Brandimonte[1]

[1]Department of Education, Psychology and Communication Sciences, Suor Orsola Benincasa University, Naples, Italy

[2]Department of Engineering, University of Cambridge, UK

Abstract

When faced with the need to transform an object, idea, or situation, people have a tendency to favor adding new components rather than removing existing ones. This is called the *additive bias*. Previous research, along with historical and anecdotal examples, shows that this bias may significantly reduce problem-solving abilities and have a detrimental impact on the innovation process. In this study, our objective was to develop a novel tool, the additive bias Implicit Association Test (ad-IAT), to investigate the reasons underlying people's preference for additive actions. By using this tool, we empirically demonstrated that people displayed an inherent tendency to assign a positive valence to additive concepts and to perceive additive actions as safer and more functional than subtractive concepts. Importantly, we also found that implicit preference for addition resulted in participants favoring additive actions while neglecting subtractive alternatives when engaged in a problem-solving task. Collectively, our series of experiments substantiated the effectiveness of our ad-IAT in uncovering and quantifying the additive bias. This, in turn, provided a deeper comprehension of the underlying factors contributing to the bias and its impact on people's behavior.

Key words: cognitive biases, additive bias, Implicit Association Test, problem solving, innovation.

Introduction

In her popular book, *The Life-Changing Magic of Tidying Up*, Maria Kondo described a common tendency for people faced with the task of organizing their homes or offices: instead of getting rid of the unused items, people often end up purchasing storage units (Kondo, 2014). With that in mind, consider the following questions: Is your mobile device overloaded with apps that you rarely use, and yet you don't delete them? Do you purchase items on sale even if you don't need them, and would be better off reducing what you own instead? If you answered yes to any of these questions, you're not alone. In striving to improve our lives, our work, and our society, we tend to add new things rather than removing things, even when such additions go against our best interests, and even when subtraction would be more effective. This tendency has recently been conceptualized as an "additive" bias (Adams, Converse, Hales, & Klotz, 2021).

If we think of our distant past, it was advantageous to accumulate resources such as food, firewood, weapons, and influence, as having more of these things increased the likelihood of survival for people, their families, and their communities. Now place this in the context of contemporary Western ideas of success and consumerism and you have an exaggeration of the additive bias. In fact, it has been suggested that this bias may contribute to various costly modern trends, including overloaded minds and schedules, bureaucratic complexities in institutions, and humanity's unsustainable encroachment on the Earth's resources (Adams et al., 2021; Frederiks, Stenner, & Hobman, 2015; Jannace & Camuffo, 2023; Kollmuss & Agyeman, 2002).

In a recent paper, Adams et al. (2021) reported eight studies that revealed the pervasive effect of the additive bias in problem solving, whether the problems were technical or social in nature. These studies crossed many fields of application, including editing a written document, modifying a mini-golf course, proposing suggestions for improving a university, and stabilizing a toy structure. In all studies, the authors observed a similar trend: participants solved problems by adding things rather than taking things away (for replication and extension see Fillon, Girandola, Bonnardel, & Souchet, 2022; Fischer, Winter, Felisatti, Myachykov, Mende, & Shaki, 2021; Jannace & Camuffo, 2023). Adams et al. (2021) suggested that the additive bias may be driven by cultural and cognitive factors. From a cultural perspective, over time, additive changes may come to be viewed more positively than subtractive changes. Numerical concepts of "more" and "higher" may align with evaluative concepts of "positive" and "better" (Gibbs, 1996; Glucksberg & McGlone, 2001; Richardson, Spivey, Edelman, & Naples, 2001). Tangible contributions are culturally valued (McAdams & de St Aubin, 1992), and acquiring and displaying resources enhances one's fitness (Stephens & Krebs, 1986; Veblen, 1994).

From a cognitive perspective, the additive bias may result from evaluative processes that lead people to favor additive actions and disregard subtractive actions. For example, people may overlook subtractive changes because of their preference for the *status quo,* defined as the tendency for maintaining one's current or previous decision (e.g. Eidelman & Crandall, 2009). From this perspective, the additive bias may be analogous to the well-known "anchoring effect" leading people to ground their decisions in the existing information that is available (the "anchor") (e.g. Tversky & Kahneman, 1974). This tendency has been extensively shown to limit the types of options that people consider when making decisions or solving problems, constraining them to the initial reference point (for a review see Furnham & Boo, 2011). Also, the human tendency to neglect subtractive choices can be considered the flip side of loss-aversion (Kahneman & Tversky, 1979), a decision-making bias which typically leads people to overemphasize the subjective value of loss outcomes relative to gain outcomes (e.g. Kahneman, Knetsch, & Thaler, 1990; Kahneman & Tversky, 1979).

Recognizing that people may "suffer" from an additive bias does not imply that subtracting is always a better option. Adding and subtracting are just complementary ways to try to change things. Difficulties arise when individuals become "fixated" on additive choices, failing to consider subtractive alternatives (Crilly & Cardoso, 2017). Research from various disciplines sheds light on the detrimental consequences of neglecting subtraction. One illustrative example can be found in the fields of computer software and manufacturing, where scholars frequently discuss the phenomenon of 'feature creep': "the excessive and continuous expansion or addition of new features in a product or service" (Sullivan, 2005). The inclusion of these unnecessary features often results in software bloat and in the creation of overly-complex products, ultimately causing consumers to experience "feature fatigue" thus having a negative impact on market performance (Thompson, Hamilton, & Rust, 2005).

The additive bias is not limited to specific domains; it permeates various aspects of everyday life. For instance, it manifests in people's inclination to buy and consume more than necessary, for example, during sales season or at *all-you-can-eat* buffets, or to continue accumulating possessions in their spaces, instead of decluttering and getting rid of unnecessary items (Kondo, 2014). Furthermore, this bias has been found to contribute to wastefulness and the excessive accumulation of resources, which ultimately hinders the adoption of more sustainable behaviors (Frederiks, Stenner, & Hobman, 2015; Kollmuss & Agyeman, 2002).

Like anything people systematically overlook, subtracting has untapped potential. With subtraction in mind, people may consider alternative and, sometimes, more functional options that would have been otherwise neglected. Indeed, many inventions and technology advances resulted from people's ability to consider subtraction when developing ideas to solve problems. For example, for many years toddlers were educated to ride bicycles by first rolling along on tricycles and then wobbling around on bikes with training wheels. In 2007, this scenario changed. Ryan McFarland found himself frustrated that the products he was using to teach his son to ride a bike were "too large, too heavy, and too complicated." His insight was that kids would learn better not on contraptions that were like bikes with added parts — namely, extra wheels for stability — but on bikes with fewer parts. After experimenting, he settled on a low-set two-wheel bike without pedals and chains. McFarland used this idea to found Strider Bikes, a US-based company dedicated to making balance bikes. Within 7 years, the company had sold almost 700,000 balance bikes; by 2017, that figure was up to 1.6 million bikes. Nowadays, many children learn the basics of balance by riding on one of McFarland's balance bikes (Klotz, 2021). Another example is Braess's paradox, discovered by the German mathematician Dietrich Braess in 1968. This paradox relates to the observation that improving a malfunctioning network can sometimes be achieved by removing certain parts of it. The paradox has been used to explain instances where adding one or more roads to a road network resulted in an unexpected increase in traffic flow (Braess, Nagurney, & Wakolbinger, 2005). Similarly, Google and Apple have become some of the world's most valuable companies through the design of minimalist graphical interfaces and minimalist physical products, often by subtracting extraneous features. In relation to this, the Apple founder Steve Jobs recognized that "innovation is saying no to a thousand things." For the business magnate, innovation lies in the ability to "turn down hundreds of good ideas and concentrate on a few carefully selected options" (Gallo, 2011).

These and other examples demonstrate that the additive bias can significantly limit the range of ideas that individuals consider when solving problems. Similar constraining effects have been observed in relation to various cognitive biases (e.g. Cristofaro, Leoni, & Giardino, 2021; Oswald & Grosjean, 2004), and have been associated with the inhibition of innovation (Johnson, 2022). For example, within the field of creativity, scholars have extensively documented the existence of numerous biases that can impact creative problem-solving. Many of these can be viewed as general biases of which the additive bias is a specific instance. For example, if we consider addition as a process that is followed then the additive bias may be related to the familiar process bias (e.g. Mueller, Melwani, & Goncalo, 2012). From this standpoint, adding may result from the tacit use of heuristics commonly employed by individuals to reduce demand, uncertainty and risk when processing information and selecting strategies to solve problems (Mumford, Blair, Dailey, Leritz, & Osburn, 2006). Alternatively, if we consider additional features to be attributes of a solution then the additive bias might be related to the familiar solution bias (e.g. Bilalić, M., McLeod, & Gobet, 2010; Neroni, Vasconcelos, & Crilly, 2017). From this other standpoint, adding might simply be a way of conforming to existing schemas or models when generating ideas to solve problems (e.g., Ward, 1994, 1995). Whether it is the underlying mechanism or the specific form of the bias, these considerations underscore the intricate relationship between the additive bias and creativity, accentuating its role in shaping decision-making and problem-solving. These observations have, in turn, prompted an expansion of studies addressing cognitive biases that undermine creativity and innovation, and have led to the development of educational interventions aimed at monitoring and mitigating their impact on people's behavior (e.g., Collino & Lauto, 2022; Liedtka, 2015).

Overall, these observations highlight the importance of understanding the additive bias, its underlying mechanisms and its influence on people's behavior. This motivates the present study, where, inspired by the principles and methods of cognitive psychology, we aimed at developing a new digital tool, employable on a large scale, that can be used to understand (and ultimately) mitigate the additive bias. In particular, the present research was organized according to the following three objectives:

1. Development (and test) of a digital tool that can reliably *detect* and *measure* the additive bias (Experiment 1 + extension),
2. Evaluation of the tool's potential to reveal the mechanisms underlying the additive bias (Experiment 2);
3. Evaluation of the tool's potential to predict people's choices and behaviors (Experiment 3).

Experiment 1

A traditional approach to test whether a new tool efficiently measures what it is supposed to measure is to evaluate the so-called "construct validity" (e.g. Greenwald, McGhee, & Schwartz, 1998). Construct validity is usually verified by comparing two experimental groups or conditions that are known *a priori* to differ with respect to the variable (or construct) of interest. For example, if a new test is designed to be a valid test of math knowledge, then math majors should outperform non-math majors on that test.

One hypothesis proposed to explain the additive bias is that people, over time, have developed associations between addition and positive concepts (Adams et al., 2021). If this is true, a valid additive bias test should reveal stronger associations between additive concepts and positive concepts than between subtractive concepts and positive concepts. But how can we measure this association? One place to look is the research on implicit attitudes and, in particular, the work done with the Implicit Association Test ('Project Implicit'; Greenwald et al., 1998). This test is a widely-used cognitive-behavioral paradigm that measures the strength of automatic associations between concepts in people's minds relying on latency measures detected in a simple sorting task (e.g. Greenwald et al., 1998; Greenwald, Nosek, & Banaji, 2003).

When completing the IAT, participants perform two different categorization tasks. In the single categorization tasks, participants classify individual stimuli into one of the two target categories (e.g. insects vs. flowers) or attributes (e.g. pleasant vs. unpleasant) by pressing two different keyboard keys (e.g. left and right key). These stages are learning stages in which participants become familiar with the categorization tasks. In the combined tasks, the previously learned categorizations are combined. Participants are instructed to press a (left) key if any given word is either an insect word or an unpleasant word and a different (right) key if any given word is either a flower word or a pleasant word. They are also instructed to press a (left) key if any given word is either a flower word or an unpleasant word and to press the other (right) key if any given word is either an insect word or a pleasant word. What is expected (and found) in this task is that people are quicker to respond when generally liked items (e.g. flowers) are paired with pleasant words than when generally disliked items (e.g. insects) are paired with pleasant words. The proposed and widely accepted explanation for these results is that it is easier for people to categorize two concepts with the same response key if they view the concepts as compatible because, in such cases, the mental processes underlying the categorization are relatively automatic and effortless. In contrast, it is more difficult for people to categorize two concepts using the same response key if the concepts are incompatible because of response competition (Greenwald et al., 1998).

Across many disciplines and contexts (e.g. Baron & Banaji, 2006; Czopp & Monteith, 2003; Egloff & Schmukle, 2002; Gemar, Segal, Sagrati, & Kennedy, 2001; Maison, Greenwald, & Bruin, 2001; Schultz, Shriver, Tabanico, & Khazian, 2004), the IAT has been extensively and reliably used to measure automatic associations. The usefulness of the IAT is due to its combination of apparent resistance to self-presentation artifacts (e.g. Banse, Seise, & Zerbes, 2001; Egloff & Schmukle, 2002), its lack of dependence on introspective access to the association strengths being measured (Greenwald et al., 2003), and its ease of adaptation to assess a broad variety of socially significant associations (Greenwald & Nosek, 2001).

Being inspired by research on the IAT, we developed an IAT variant (i.e., the additive bias IAT, ad-IAT) that could be used to detect and measure a person's level of mental association between additive concepts and pleasant words and/or between subtractive concepts and unpleasant words. If, as expected, over time, socio-cultural factors or other reasons have led people to associate additive concepts with positive concepts (Adams et al., 2021), participants should respond faster and, possibly, more accurately when additive concepts are paired with pleasant words and subtractive concepts are paired with unpleasant words, *as compared* to when additive concepts are paired with unpleasant words and subtractive concepts are paired with pleasant words.

Method

Participants

One-hundred and fifty participants were initially recruited into the study by responding to a posted advertisement published on the online platform Amazon Mechanical Turk (for methodological considerations, see Highhouse & Zhang, 2015). We restricted the recruitment to participants who were native English speakers and aged 18 years or older. From the original sample, we removed 23 participants as they either did not complete the study (7 participants) or they met the exclusion criteria as recommend by Greenwald et al. (2003)[1] (16 participants). The final sample therefore consisted of 126 participants (52 females).

The participants' average age was 32.07 years (SD = 5.07). Most of the participants had a bachelor's degree (49.61%) as their highest qualification, followed by those having a master's degree (31.71%). Only a few participants had a college qualification (7.09%) or a high school diploma (4.72%) as their highest qualifications. Participants were predominantly from North America (67.72%), followed by those from Europe (15.75%). Only a few participants were from South America (9.45%), Asia (2.63%), Australia (1.57%) and Africa (1.57%).

Before starting with the study, participants read an on-screen information screen and gave their consent to participate. Participants received a small honorarium (5 USD) in return for their participation. The entire study procedure adhered to the ethical principles recommended by the American Psychological Association (https://www.apa.org/ethics/code).

Materials

The IAT materials consisted of two sets of words, organized into a target category and into an attribute category. The target category comprised verbs associated with additive actions (e.g., increase, raise, enlarge) and subtraction actions (e.g., remove, decrease, reduce).

To obtain these sets of verbs, a pre-testing session was conducted, involving a sample of fifty individuals aged between 18 and 40 years. They were informally asked to categorize 30 English verbs as either "add" or "subtract." These verbs were selected from online dictionaries, with half of the verbs related to addition actions and the other half related to subtraction actions. Based on the feedback received, 10 "addition" verbs and 10 "subtraction" verbs were chosen, prioritizing those with the highest consensus among the participants.

The attribute category consisted of 10 pleasant and 10 unpleasant words, which were selected from previous IAT studies (Greenwald et al., 1998). The full set of words used to build the IAT is reported in Table 1.

[1] In line with Greenwald et al. (2003, p. 214, Table 4), participants were excluded if more than 10% of all their response latencies were faster than 300ms.

Table 1. Categories and related words.

Categories	Category types	Stimuli
Add	Target category	Increase, Augment, Raise, Enlarge, Expand, Extend, Heighten, Intensify, Amplify, Widen.
Subtract	Target category	Decrease, Reduce, Lower, Diminish, Lessen, Eliminate, Remove, Cancel, Shorten, Delete.
Pleasant	Attribute category	Joyous, Nice, Delight, Fabulous, Magnificent, Rainbow, Smile, Joy, Sunshine, Lovely.
Unpleasant	Attribute category	Sad, Painful, Horrible, Hurtful, Disgusting, Failure, Poison, Vomit, Sickness, Death.

Procedure

The IAT was developed, hosted and administered through the Inquisit Web platform (https://www.millisecond.com/). When accessing the link for the study, participants were immediately required to download the Inquisit app to their computers. Once the Inquisit app was downloaded, participants were asked to declare that they were both older than 18 years and native-English speakers. Participants were also required to run the study on a computer equipped with a keyboard (not on a tablet or on a smartphone). Participants then viewed preliminary information that described what they might experience in taking an IAT, and were offered the opportunity to continue if they wished to do so. If participants continued with the study, they were presented with the general instructions for the IAT and proceeded to complete it.

During the IAT, participants were asked to categorize stimuli (target and attributes categories) into different categories by pressing a left key ("I") if the word belonged to the category presented on the left (e.g. "ADD") and to press the right key ("E") if the word belonged to the category presented on the right (e.g. "SUBTRACT"). This assignment is based on key locations on a QWERTY keyboard.

Specifically, the IAT consisted of seven blocks (see Table 2). Each block began with an instruction screen that provided participants with detailed information about the tasks they were required to perform in that specific block. In the single categorization blocks (Blocks 1, 2 and 5), participants had to press a key if the stimulus belonged to *one* target/attributes category (e.g. "ADD", "pleasant") and to press the other key if the stimulus belonged to the opposite target/attributes category (e.g. "SUBTRACT", "unpleasant"). In the combined categorization blocks, the two sets of stimuli (i.e. target and attributes category stimuli) were combined. Specifically, in the combined compatible blocks (Blocks 3 and 4), participants had to press one key for "ADD" and "pleasant" and a different key for "SUBTRACT" and "unpleasant". In the combined incompatible blocks (Blocks 6 and 7), participants had to press one key for "ADD" and "unpleasant" and a different key for "SUBTRACT" and "pleasant".

Table 2. Addition bias IAT structure.

Block	Type of task	Left key assignment	Right key assignment	Number of trials
B1	Single categorization	Add	Subtract	20
B2	Single categorization	Pleasant	Unpleasant	20
B3	Combined compatible	Add Pleasant	Subtract Unpleasant	20
B4	Combined compatible	Add Pleasant	Subtract Unpleasant	40
B5	Single (target) reversed categorization	Subtract	Add	20
B6	Combined incompatible	Subtract Pleasant	Add Unpleasant	20
B7	Combined incompatible	Subtract Pleasant	Add Unpleasant	40

Target and attribute stimuli appeared in random order across the blocks. However, as recommended by Gawronski (2002), within each block, the stimuli appeared in the same order for all participants in an effort to control for general confounding from individual differences. Also, in line with consolidated IAT procedures, in the combined blocks a target category stimulus was never followed by another target category stimulus, instead it was always followed by an attribute category stimulus (or vice versa). Importantly, we deviated from conventional IAT procedures (e.g., Greenwald et al., 1998) by implementing a comprehensive counterbalancing approach for the left/right key assignments. This involved dividing participants into two groups: one group completed the task with inverted target labels (from Block 5 to Block 7), while the other group completed the task with inverted attribute labels (from Block 5 to Block 7). This counterbalancing procedure was implemented to control for potential confounding effects associated with the associations between adding/subtracting concepts and spatial positions (e.g., Richardson, Spivey, Edelman, & Naples, 2001).

Each stimulus appeared once in the practice combined blocks (i.e. Block 3 and 6, see Table 2) and twice in all the other blocks. In all the blocks, stimuli appeared vertically and horizontally centered on the screen against a black background. Target category stimuli and corresponding labels appeared in light-blue uppercase letters; attributes category stimuli and corresponding labels appeared in light-green lowercase letters. Target and attributes labels were positioned above the stimuli, on the top left and top right side on the screen, and always remained in view to remind participants of the response key mapping rules. These reminder labels were “ADD” and “SUBTRACT” for single target-classification blocks, “pleasant” and “unpleasant” for single attribute-classification blocks. Combined blocks also showed appropriate labels (e.g., “ADD *or* pleasant”, “SUBTRACT *or* unpleasant”).

While completing the task, participants were required to give their responses as quickly and accurately as possible. If they made a mistake, a red “X” appeared, and participants had to fix the error (by hitting the correct response key) to move on with the task. In line with Greenwald et al.’s IAT procedure (1998, 2003), participants were given unlimited time to make a response.

After completing the IAT, we collected general demographic information related to the participants’ gender, age, education and nationality. In addition, participants were required to indicate their level of prior experience with IATs on a 7-point scale (i.e. 1 = “I have never completed this or similar tasks”; 7 = “I have completed this or similar tasks several times”). This question was included to account for the observation made by Nosek, Banaji and Greenwald (2002) that prior experience with IATs can influence subsequent IAT performance.

The whole session lasted about 20 minutes on average.

Results

The IAT data were processed using the D-score data-cleaning and scoring algorithm (for a full description see Greenwald et al., 2003). This algorithm was found to be the one that worked best to minimize (1) the correlation between IAT effects and individual subjects’ average response latencies, (2) the effect of the order of the IAT blocks, and (3) the effect of previously completing one or more IATs on IAT scores, while (4) retaining strong internal consistency and (5) maximizing the correlation between implicit and explicit measures (Lane, Banaji, Nosek, & Greenwald, 2007).

Briefly, for each participant, a D-score was computed by calculating the difference in average response latency (i.e. the latency of the final, correct response in ms, measured from onset of the word) between the IAT’s combined blocks, divided by an “inclusive” standard deviation of the participant’s response latencies in the combined blocks. In line with this recommended algorithm, only latencies that were less than 10s were considered for D-score calculations.

D-scores provide information about the direction as well as the strength of the association. In this Experiment, positive D-scores indicated that the association between *additive* concepts and pleasant words and between subtractive concepts and unpleasant words was stronger *than* the association between additive words and unpleasant words and between subtractive words and pleasant words. This was assumed to be our IAT effect (Greenwald et al., 1998, 2003).

To assess the internal consistency of the IAT, we computed the correlation between a D-score derived from the practice blocks (Blocks 3 and 6) and a D-score based on the critical blocks (Blocks 4 and 7), following the methodology outlined by Greenwald et al. (2003). We found a significant positive correlation between the D-scores for the practice blocks and the D-scores for the critical blocks (r = 0.59, p < .001). Given this outcome, we proceeded to conduct all subsequent analyses using an overall D-score, which was obtained by calculating the unweighted mean of the two D-scores.

Overall, the results revealed a strong IAT effect (*d* = 0.67).[2] Analyses of responses times revealed that participants were faster at categorizing stimuli in the compatible blocks (mean response times = 1002.50 ms, *SD* = 415.97) than in the incompatible blocks (mean response times = 1477.29 ms, *SD* = 718.56) than, *t* (125) = -12.31, $p < .001$. Also, participants' accuracy rates (i.e. number of correct responses) were higher in the compatible blocks (87%) than in the incompatible blocks (77%*), χ2* = 2334.18, $p < .001$. Analysis of the participants' responses in the post-experiment interview indicated that the majority of participants had limited prior experience with the task (from 1 to 7, *M* = 2.06, *SD* = 1.03).

Experiment 1: extension

Extensive research in linguistics, anthropology, and psychology has revealed substantial variations in the meanings of words across languages (Evans & Levinson, 2009; Majid, Bowerman, Kita, Haun, & Levinson, 2004). These linguistic differences have been found to have a significant impact on cognition (e.g. Lupyan, Rahman, Boroditsky, & Clark, 2020; Richardson et al., 2001). For instance, the linguistic relativity hypothesis (Wolff & Holmes, 2011) suggests that the structure and vocabulary of one's native language can influence how individuals conceptualize and perceive the world. These observations underscored the importance of ensuring that the effectiveness of our word-based IAT in detecting and measuring the additive bias was not significantly undermined by linguistic disparities.

To evaluate the generalizability of our tool in investigating the additive bias across different cultures and languages, we developed (and tested) an Italian version of our ad-IAT. This involved a meticulous process of translating the test materials (as described in the Method section of Experiment 1) from English into Italian. Specifically, by employing an Italian version of the task, we aimed at ascertaining whether the findings from Experiment 1 would be replicated. We anticipated observing similar results, whereby participants would exhibit faster response times and higher accuracy when categorizing "add" and "pleasant" using the same response key, and "subtract" and "unpleasant" using the same response key, compared to categorizing "add" with "unpleasant" and "subtract" with "pleasant."

Method

Participants

Sixty-eight people were invited to participate in this experiment through a university-based recruitment pool. We restricted the recruitment to people who were native Italian speakers and aged 18 years or older. From the original sample, we removed 18 participants as they did not meet the exclusion criteria as recommend by Greenwald et al. (2003). The final sample therefore consisted of 50 participants (41 females). Their average age was 22.48 years (SD = 1.30). All the participants had a high school diploma as their highest qualification.

---

[2] D-scores suggested interpretation (Greenwald et al., 2003): D-score $\leq$ 0.15 indicates a "slight" preference for hypothesis-confirming pairings; D-score <= 0.35 indicates a "moderate" preference for hypothesis-confirming pairings; D-score $\leq$ 0.65 indicates a "strong" preference for hypothesis-confirming pairings.

Before starting with the study, participants read an on-screen information sheet and gave their consent to participate. Participants did not receive any reward in return for their participation. The entire study procedure adhered to the ethical principles recommended by the American Psychological Association (https://www.apa.org/ethics/code).

Materials and Procedure

As in Experiment 1, the IAT stimuli consisted of two sets of words, organized into a target category and into an attribute category. The target category included Italian translations of English verbs related to additive actions and to subtractive actions (as listed in Table 1). The attribute category included Italian translations of pleasant and unpleasant words, as used in Experiment 1, selected from those used in previous IAT studies (Greenwald et al., 1998).

The full set of words used to build the Italian version of the ad-IAT is reported in Table 3.

Table 3. Categories and related words (in Italian).

Categories	Categories types	Stimuli
Aggiungere	Target category	Aumentare, Incrementare, Addizionare, Allargare, Espandere, Estendere, Accrescere, Annettere, Amplificare, Ampliare.
Sottrarre	Target category	Decrescere, Ridurre, Diminuire, Togliere, Levare, Eliminare, Rimuovere, Cancellare, Accorciare, Cancellare.
Piacevole	Attribute	Gioioso, Carino, Delizioso, Favoloso, Magnifico, Arcobaleno, Sorriso, Gioia, Sole, Amore.
Spiacevole	Attribute	Triste, Doloroso, Orrendo, Doloroso, Disgustoso, Fallimento, Morte, Vomito, Malattia, Veleno.

As with Experiment 1, the IAT was developed, hosted and administered through the Inquisit Web platform. Recruited participants were sent the link for the study and were required to run the experiment on a computer equipped with a keyboard (not on a tablet or a smartphone). Once the Inquisit app was downloaded, participants were asked to declare that they were older than 18 years and were native-Italian speakers.

The entire task procedure was identical to that described in Experiment 1. Following completion of the task, we collected general demographic information from the participants, along with data regarding their previous experience with IATs.

Results

As in Experiment 1, the IAT data were processed using the D-score data-cleaning and scoring algorithm (for a full description see Greenwald et al., 2003).

Internal IAT consistency was calculated as the correlation between a D-score based on the practice blocks (Blocks 3 and 6) and a D-score based on the critical blocks (Blocks 4 and 7; see Greenwald et al., 2003). Correlation analysis revealed the existence of a positive correlation between d-score for the practice blocks and d-scores for the critical blocks, $r = 0.50$, $p < .001$. In light of this result, we conducted all the analyses on an overall d-score, calculated as a non-weighted mean of the two d-scores.

The results revealed a strong IAT effect (*d* = 0.88). Analyses of response times revealed that participants were faster at categorizing stimuli in the compatible blocks (mean response times = 951.51, *SD* = 249.64) than in the incompatible blocks (mean response times = 1620.45, *SD* = 462.77) than, t (49) = -11.18, $p < .001$. Also, participants' accuracy rates (i.e. number of correct responses) were higher in the compatible blocks (96%) than in the incompatible blocks (91%), $\chi 2$ = 176.93, $p < .05$. During the post-experimental interview, participants indicated a limited previous experience with the task (*M* = 1.30, *SD* = 0.46).

Discussion

In Experiment 1, and its extension, we were inspired by research on the IAT to develop an additive bias IAT that could reveal and measure people's preference for adding concepts. The results were consistent with our hypotheses. Specifically, we found that participants were faster and more accurate (i.e. correct key assignments) to respond when "add" and "pleasant" shared the same response key and "subtract" and "unpleasant" shared the same response key, *compared to when* "add" and "unpleasant" shared the same response key and "subtract" and "pleasant" shared the same response key. This result provides the first empirical evidence to support Adams et al.'s speculation (2021) that, "over time, additive changes may come to be viewed more positively than subtractive changes" (p. 259).

While we can speculate on the underlying mechanisms behind people's preference for additive concepts, in Experiment 1 our objective was purely methodological: developing a tool capable of uncovering and quantifying an individual's preference for adding actions. The results from Experiment 1 showed that we successfully achieved this objective. Importantly, our observation that people's preference for additive concepts, as detected by our IAT, is English-independent highlights the potential for employing the task, through translation, to explore the additive bias across cultures.

In developing our ad-IAT, we focus on the attributes dimension most commonly and reliably used in previous IAT studies (e.g. Schnabel, Asendorpf, & Greenwald, 2008). However, it is possible that the pleasant vs. unpleasant evaluative dimension is not the only one that underlies people's preference for additive concepts. For example, people may prefer additive actions because they feel they are safer or more functional, and downgrade subtractive actions because they feel they are riskier or less functional.

In light of these considerations, in Experiment 2 we aimed at more deeply investigating the types and natures of the mental associations underlying the additive bias. To do so, we developed an extended version of our ad-IAT, in which additive and subtractive concepts were paired with multiple (all potentially relevant) attribute dimensions.

Experiment 2

Aiming at more deeply exploring the nature of the associative mechanisms underlying the additive bias, we developed an extended, multidimensional version of our "additive bias" IAT, the 'md-IAT' (Gattol, Sääksjärvi, & Carbon, 2011). Specifically, instead of employing just one IAT and therefore only one attribute dimension (i.e. pleasant–unpleasant), we used three consecutive IATs, each intended to measure associations between our target concepts (i.e. additive and subtractive concepts) and a particular attribute dimension.

The attribute dimensions were selected on the basis of some of the cognitive mechanisms hypothesized to underly the additive bias, namely (1) an (implicit) preference for additive concepts (e.g. Adams et al., 2021; our Experiment 1), (2) a possible perceived higher functionality of additive strategies (e.g. Higgins, 1996; Wiley, 1998), and (3) an overall human tendency to favour the *status quo* (Eidelman et al., 2009) and avoid losses (e.g. Kahneman & Tversky, 1979). Consequently, the bipolar attribute dimensions used to build the mad-IAT were: (1) pleasant – unpleasant, (2) functional – unfunctional, (3) safe – unsafe.

In previous research, the md-IAT has been used to study multidimensional implicit processes underlying different types of human behaviors, including consumers' attitudes towards products (Gattol et al., 2011), people's attitudes towards visual stimuli (Bertamini, Makin, & Rampone, 2013) and implicit mental associations underlying stereotypes (Brandenstein, Gebauer, & Carbon, 2019). The md-IAT was also tested regarding its psychometric criteria and it was shown to be a reliable, valid and sensitive indirect measure of associations (Gattol et al., 2011).

If the additive bias reflects automatic evaluative associations that lead people to favor additive options over subtractive options (Adams et al., 2021), participants should be faster and make less errors in the task blocks where the category "add" and the attributes related to the dimensions "pleasant", "functional" and "safe" share the same response (and in the blocks where "subtract" and the attributes related to the dimensions "unpleasant", "unfunctional", "risky" share the other response), compared to the blocks in which "add" and the attributes related to the dimensions "unpleasant", "unfunctional", "risky" share one response (and in the blocks where "subtract" and the attributes related to the dimensions "pleasant", "functional" and "safe" share the other response).

Method

Participants

Eighty participants were initially recruited into the study by responding to a posted advertisement published on the online platform Amazon Mechanical Turk. We restricted the recruitment to participants who were native English speakers and aged 18 years or older. From the original sample, we removed 31 participants as they either had not completed the study (4 participants) or they met the exclusion criteria in at least one dimension of the IAT as recommend by Greenwald et al. (2003) (27 participants). The final sample consisted therefore of 49 participants (22 females).

Participants' average age was 33.69 years (SD = 4.93). Most of the participants had a bachelor's degree (38.77%) or a master's degree (38.77%) as their highest qualifications. Only a few participants had some college qualification but not a degree (10.20%) or a high school diploma (10.20%) as their highest qualifications. Participants were predominantly from North America (69.39%), followed by those from South America (12.24%) and from Europe (10.20%). Only a few participants were from Asia (4.08%), Australia (2.04%) and Africa (2.04%).

Before starting the study, participants read an information screen and gave their consent to continue. Participants received a small honorarium (5 USD) in return for their time. The entire study procedure adhered to the ethical principles recommended by the American Psychological Association (https://www.apa.org/ethics/code).

Materials and Procedure

In the mad-IAT, the target category comprised a set of verbs associated with addition and a set of verbs associated with subtraction, which were pre-tested and used in Experiment 1 (refer to Table 1 for details). The attribute category varied across the different IATs employed in the study. Specifically, the pleasant and unpleasant attributes were those used in Experiment 1 and selected from previous IAT studies (e.g. Greenwald et al., 1998).

Regarding the other two attribute sets, stimuli were selected following a similar procedure to that outlined in the Method section of Experiment 1. Initially, a series of synonyms for functional, nonfunctional, safe, and unsafe were identified from online dictionaries. A total of 12 synonyms were chosen for each attribute. Subsequently, a group of 10 native English speakers were informally asked to rate the extent to which they perceived each word as representing the corresponding reference attribute, on a 7-point scale (i.e. 1 = the word has exactly the same meaning as the reference attribute; 7 = the word has a completely different meaning from the reference attribute). Based on the ratings provided by the participants, the 10 words with the highest scores for each attribute were selected. The complete set of words used to construct the mad-IAT is presented in Table 4.

Table 4. Categories and corresponding words used to build the mad-IAT.

Categories	**Category types**	**IAT**	**Stimuli**
Add	Target category	1, 2, 3	Increase, Augment, Raise, Enlarge, Expand, Extend, Heighten, Intensify, Amplify, Widen.
Subtract	Target category	1, 2, 3	Decrease, Reduce, Lower, Diminish, Lessen, Eliminate, Remove, Cancel, Shorten, Delete.
Pleasant	Attribute	1	Joyous, Nice, Delight, Fabulous, Magnificent, Rainbow, Smile, Joy, Sunshine, Lovely.
Unplesant	Attribute	1	Sad, Painful, Horrible, Hurtful, Disgusting, Failure, Poison, Vomit, Sickness, Death.
Functional	Attribute	2	Practical, Useful, Useful, Operative, Utilitarian, Functioning, Applicable, Workable, Productive, Effective, Feasible.
Nonfunctional	Attribute	2	Defective, Faulty, Flawed, Fruitless, Futile, Malfunctioning, Useless, Weak, Abortive, Disadvantageous.
Safe	Attribute	3	Secure, Intact, Trusted, Steady, Innocuous, Certain, Defended, Reliable, Guarded, Sheltered.
Unsafe	Attribute	3	Dangerous, Hazardous, Risky, Precarious, Perilous, Vulnerable, Fragile, Tricky, Speculative, Threatened.

Overall, the IAT procedure was the same as in Experiment 1. Each IAT consisted of seven blocks: B1 and B2 (single categorization tasks), B3 and B4 (combined compatible categorization tasks), B5 (reversed single categorization task), B6 and B7 (combined incompatible categorization tasks) (see Table 2). Specifically, in the "pleasant/unpleasant" IAT (referred to as IAT1), participants sorted add/subtract stimuli along with pleasant/unpleasant stimuli. In the "functional/nonfunctional" IAT (referred to as IAT2), participants sorted add/subtract stimuli along with functional/unfunctional stimuli. In the "safe/unsafe" IAT (referred to as IAT3), participants sorted add/subtract stimuli along with safe/unsafe stimuli.

The full task procedure was the same as the one outlined in Experiment 1. The order of the three IATs was fully counterbalanced across participants. Also, as in Experiment 1, across the full mad-IAT, we counterbalanced the left/right key assignments by having one group of participants completing the three IATs with inverted target labels (from Blocks 5 to Blocks 7) and the other group of participants completing the IATs with inverted attributes labels (from Blocks 5 to Blocks 7).

After completing all the IATs, as in Experiment 1, we also collected general demographic information and asked participants to indicate their level of prior experience with IATs on a 7-point scale (i.e. 1 = "I have never completed this or similar tasks"; 7 = "I have completed this or similar tasks several times").

Results

As in Experiment 1, data preparation of the IATs data was based on the improved algorithm criteria by Greenwald et al. (2003) and included adjustments depending on reaction times, as well as D-scores calculation.

In Experiment 2:

- positive D-scores in IAT1 indicated that the association between additive concepts and pleasant words and between subtractive concepts and unpleasant words is stronger *than* the association between additive concepts and unpleasant words and between subtractive concepts and pleasant words;
- positive D-scores in IAT2 indicated that the association between additive concepts and functional words and between subtractive concepts and non-functional words is stronger *than* the association between additive concepts and non-functional words and between subtractive concepts and functional words;
- positive D-scores in IAT3 indicated that the association between additive concepts and safe words and between subtractive concepts and unsafe words is stronger *than* the association between additive concepts and unsafe words and between subtractive concepts and safe words.

As in Experiment 1, the internal consistency of the IATs was initially calculated as the correlation between a D-score based on the practice blocks (Blocks 3 and 6) and a D-score based on the critical blocks (Blocks 4 and 7). For all the IATs, the results revealed a positive correlation between D-scores for the practice blocks and D-scores for the critical blocks (IAT1: $r = 0.66$, $p <. 001$; IAT2: $r = 0.51$, $p <. 001$; IAT3: $r = 0.55$, $p <. 001$). In light of this result, we conducted all the analyses on each IAT's overall D-score, calculated as a non-weighted mean of the two D-scores.

D-scores, mean response times and percentage of correct responses are presented in Table 5.

Table 5. Mean response times, percentage of correct responses and t-test values for compatible and incompatible blocks, in IAT 1, IAT 2 and IAT3.

	D-score	Mean Response Times (milliseconds)			Percentage of correct responses		
		Compatible blocks	Incompatible blocks	*T*-values	Compatible blocks	Incompatible blocks	χ2-values
IAT 1	0.77	975.69 (*SD* = 356.47)	1433.12 (*SD* = 536.05)	*t* (48) = -10.06, *p* < .001	93%	85%	*χ2* = 450.33, *p* < .001
IAT 2	0.55	959.35 (*SD* = 333.04)	1257.22 (*SD* = 395.43)	*t* (48) = -10.02, *p* < .001	87%	81%	*χ2* = 576.43, *p* < .001
IAT 3	0.43	1080.57 (*SD* = 398.96)	1311.50 (*SD* = 442.75)	*t* (48) = -5.70, *p* < .001	87%	83%	*χ2* = 472.15, *p* < .01

As shown in Table 5, D-scores ranged from *d* = 0.43 (IAT3) to *d* = 0.77 (IAT1), with all *T* values above 6.79 and all *p*-values below .001. For all the IATs, analyses of response times revealed that participants were faster at categorizing stimuli in the compatible blocks than in the incompatible blocks (all *t*s above –5.70 and all *p*-values below .001, see Table 5). Also, in all the IATs, participants accuracy rates (i.e. percentages of correct responses) were higher in the compatible blocks than in the incompatible blocks (all *χ2* above 450.33 and all *p*-values below .01, see Table 5). Analysis of the participants' responses in the post-experimental interview indicated that they had limited prior experience with the task (*M* = 2.10, *SD* = 1.10).

Discussion

In Experiment 2, we aimed to delve deeper into the nature and specific associations underlying the additive bias. Inspired by variations in IAT methodology used in previous studies (e.g., Brandenstein et al., 2019; Gattoll et al., 2011), we developed an extended version of our “additive bias” IAT, called the mad-IAT (multidimensional IAT). This mad-IAT consisted of three consecutive IATs, each based on a different bipolar attribute dimension: (1) pleasant - unpleasant (IAT1), (2) functional - unfunctional (IAT2), and (3) safe - unsafe (IAT3).

The results from IAT1 replicated the findings from Experiment 1. Specifically, we observed that participants exhibited faster response times and higher accuracy when categorizing “add” and “pleasant” using the same response key, and “subtract” and “unpleasant” using the same response key, *compared to* categorizing “add” with “unpleasant” and “subtract” with “pleasant”.

Results from IAT2 showed that participants exhibit faster response times and higher accuracy when categorizing “add” and “functional” using the same response key, and “subtract” and “non-functional” using the same response key, *compared to* categorizing “add” with “non-functional” and “subtract” with “functional”.

Finally, results from IAT3 showed that participants exhibit faster response times and higher accuracy when categorizing “add” and “safe” using the same response key, and “subtract” and “unsafe” using the same response key, *compared to* categorizing “add” with “unsafe” and “subtract” with “safe”.

Overall, these results provide a broader understanding of the automatic associations underlying the additive bias. As mentioned in the Introduction, aside from cultural and socio-ecological factors, other cognitive mechanisms may contribute to people's tendency to prefer additive actions. For example, the additive bias may be linked to the "anchoring effect" that typically leads people to ground their decisions on the initial reference point (the "anchor") (e.g. Tversky & Kahneman, 1974). People could assume that existing items or features are there for a reason, and so also assume that looking for additions would be more effective (Meyvis & Yoon, 2021). People's preference for additive actions might additionally be explained by the general human tendency to be loss-averse (e.g. Kahneman & Tversky, 1979; Kahneman et al., 1990), being more sensitive to losses than gains. Research within the domain of risk management supports this view, revealing a preference for actions that allow flexibility when confronted with problems admitting multiple solutions. Such an approach enables individuals to modify decisions and avoid committing to irreversible solutions or approaches (Gilbert & Ebert, 2002; Lovallo & Kahneman, 2000). Viewed from this perspective, the act of adding may be considered one of the strategies individuals employ to effectively manage and mitigate risk during the problem-solving process. These considerations might explain why people instinctively perceive additive actions as more functional and safer compared to subtractive actions, as evidenced in Experiment 2 for IAT2 and IAT3.

In summary, the results from Experiment 2 further validate the effectiveness of the ad-IAT as a reliable tool for detecting and quantifying the additive bias. Furthermore, these findings shed light on the underlying factors contributing to individuals' preference for additive actions.

Considering the main methodological objective of this research, employing a multidimensional IAT (mad-IAT), instead of a single IAT, offers the advantage of providing a more detailed assessment of implicit attitudes towards addition and subtraction. This may allow researchers to identify and define more nuanced profiles, potentially opening up opportunities for educational interventions aimed at addressing the bias. Therefore, the next phase of this research aimed to examine the potential impact of implicit attitudes towards additive concepts, as measured by our ad-IAT, on individuals' performance in problem-solving tasks involving both additive and subtractive options.

Experiment 3

Previous research provided compelling evidence that the IAT not only uncovers implicit attitudes and preferences but also has the ability to predict behavior. A comprehensive meta-analysis conducted by Greenwald, Poehlman, Uhlmann, and Banaji (2009) revealed significant associations between the IAT and various behavioral indicators. These indicators encompassed a wide range of behaviors, including self-reported measures, judgments, and choices. The findings underscored the IAT's capacity to predict diverse cognitive phenomena, including nonverbal behaviors, impression formation, shyness, anxiety, consumer choices and even voting patterns.

For instance, some studies revealed that individuals with more negative attitudes towards Black people (in comparison to White people) tend to make more negative evaluations of ambiguous actions performed by Black people (Rudman, Feinberg, & Fairchild, 2002; similar findings were observed in a study involving a Turkish target by Gawronski, Geschke, & Banse, 2003). Moreover, these negative attitudes were associated with unfavorable behaviors during interactions with Black individuals, such as reduced speaking time and fewer instances of smiling (McConnell & Leibold, 2001). Similarly, individuals who exhibited more negative attitudes towards people with AIDS also showed spontaneous avoidance tendencies towards these individuals (compared to healthy individuals, Neumann, Hulsenbeck, & Seibt, 2004). Furthermore, positive attitudes towards alcohol (as detected by the IAT) correlated with the extent to which heavy drinkers displayed arousal and a desire to consume alcohol when exposed to a glass of beer (Palfai & Ostafin, 2003).

In addition to these macro-level behaviors, the IAT was also found to predict lower-level perceptual and cognitive phenomena. For example, in a series of compelling experiments, Hugenberg and Bodenhausen (2003) found that people who exhibited negativity towards Black individuals in the IAT exhibited a lower threshold for detecting hostility in Black faces, but not in White faces (Hugenberg & Bodenhausen, 2003).

Overall, these findings suggest that implicit attitudes towards additive and subtractive concepts, as assessed by our ad-IAT, could potentially impact an individual's decision-making and problem-solving abilities. To test this hypothesis, in Experiment 3, we investigated the relationship between participants' scores on the ad-IAT and their performance in a problem-solving task that required employing either additive or subtractive solution approaches.

In this Experiment, we employed the "pleasant vs. unpleasant" IAT (referred to as IAT1 in Experiment 2) for two primary reasons: first, it was based on the most validated attributes stimuli (Greenwald et al., 2003); second, it exhibited the strongest IAT effect, as indicated in the Results section of Experiment 2. In light of these considerations, we believed that this version of the IAT could more accurately and effectively detect preferences for additive concepts.

For the problem-solving task, we used the grids task originally developed and employed by Adams et al. (2021) in their studies. We chose this task because of its adaptability for online administration and because, to our knowledge, it is the only available and validated task that can empirically test preferences for additive choices over subtraction choices (Adams et al., 2021, Experiments 5-8). The grids task requires participants to create perfectly symmetrical 10×10 colored grids from top to bottom and left to right by toggling the cells (changing their colors) using the fewest actions possible. In their study, Adams et al. found that a considerable number of participants solved the grids by adding colored cells instead of subtracting them, even though this approach required more mouse clicks.

In our third experiment, all participants first completed four grids, which were the same as those used by Adams et al. (2021) in their experiments (Experiments 5-8), and then proceeded to complete the ad-IAT.

If preferences for additive concepts, as measured by our ad-IAT, can predict the choice and use of additive strategies during problem-solving, we would expect to observe a negative correlation between a participant's IAT score and their frequency of using subtractive transformations when solving the grids (i.e. participants with higher D-scores would be less likely to adopt subtractive transformations).

Methods

Participants

Fifty participants (22 female) were recruited into the study by responding to a posted advertisement published on the online platform Amazon Mechanical Turk. We restricted the recruitment to participants who were native English speakers and aged 18 years or older.

The participants' average age was 32.50 years (SD = 7.88). Most of the participants had a master's degree (68%) as their highest qualification, followed by those having a bachelor's degree (32%). Only a few participants had a high school diploma (6%) as their highest qualification. Participants were predominantly from North America (68%), followed by those from South America (22%) . Only a few participants were from Europe (6%) and Asia (4%).

Before starting the study, participants read an information screen and gave their consent to participate. Participants received a small honorarium (5 USD) in return for their participation. The entire study procedure adhered to the ethical principles recommended by the American Psychological Association (https://www.apa.org/ethics/code).

Materials and Procedure

At the beginning of the Experiment, participants were informed that they would complete two computer-based tasks. Then, all the participants were presented with the Adams et al.'s grids task. Initial instructions explained that the task required the participants to transform a series of digital grid patterns. Then, participants entered a familiarization phase to learn how to interact with the workspace. On this familiarization screen, participants could interact with a 10×10 grid that functioned in the same way as the grids in the subsequent trials. Half of the blocks for the grid on the orientation screen were already green (the top left and bottom right quadrants), and the other blocks were white (top right and bottom left quadrants). Participants could click any one of the squares to see how it would change the color.

Once the familiarization phase was completed (there was no time limit to complete this phase), participants moved to a different screen and received specific instructions for the task. In particular, they learned that the objective of the task was to change the forthcoming grids patterns to make them perfectly symmetrical from left to right, and from top to bottom. Importantly, participants were also required to reach this objective by using the fewest possible mouse clicks. The phrase 'using the fewest possible mouse clicks' was displayed in bold font.

In this Experiment, we used the same grids as those used by Adams et al. (2021, Experiments 5-8). Importantly, these grids were designed so that making them symmetrical with additive transformations would require more clicks than would making them symmetrical with subtractive transformations (see Figure 1). Participants could therefore not achieve the correct response without thinking of subtractive actions.

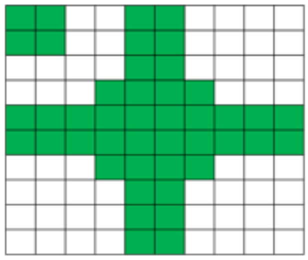
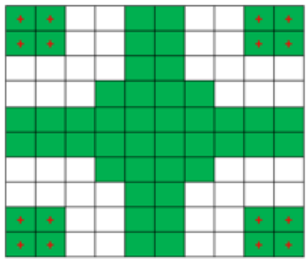
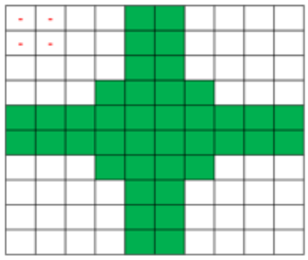
The challenge Make the grid pattern symmetrical using the fewest clicks possible. Clicking any white square will turn it green, and clicking any green square will turn it white.	
Example of additive solution: Adding green boxes to the three empty corners using 12 clicks.	
Subtractive solution: Removing the top-left corner green squares in 4 clicks.	

Figure 1. Example of a grid used in Adams et al.'s (2021) study with description of additive and subtracting solution options.

All the participants completed the four grids in a fixed order. Specifically, they worked on each grid and once they were satisfied with the solution pattern they obtained, they submitted their response and moved on to the next grid. There was no time limit for submitting each grid. During the task, participants did not receive any feedback on their performance.

After completing the grids task, participants moved to the ad-IAT. The materials and procedure for the IAT remained consistent with those outlined in Experiment 1, including the use of the same stimuli (refer to Table 1 for details). As in the previous experiments, demographic information and data pertaining to participants' previous experience with the IAT were collected upon the completion of both tasks.

Results

The results showed a significant negative correlation between participants' IAT D-scores and the number of grids solved using subtractive transformations (r = -0.86, p < 0.001). This indicated that individuals with stronger implicit attitudes towards additive concepts tended to use fewer subtractive transformations.

As Adams et al.'s (2021) observed, a difference was found in the participants' performance (resulting in an increased use of subtracting transformations) between the first three grids (referred to as "practice grids") and the last grid (referred to as "main grid"). Further analyses were conducted separately for the first three "practice grids" and the fourth "main grid". Consistent with our observations for the overall participants' performance, the results demonstrated a negative correlation between participants' IAT D-scores and the frequency of use of subtractive transformations when solving both the "practice grids" ($r = -.79$, $p < .001$) and the "main grid" ($r = -.75$, $p < .001$).

It is worth noting that only a small proportion of participants (10%) successfully solved the grids using subtractive transformations across all four grids. However, there was an increase in this percentage as participants progressed from the practice grids (10% using subtracting transformations) to the main grid (18% using subtracting transformations), which aligns with the findings of Adams et al. (2021).

During the post-experimental interview, participants indicated a limited previous experience with the task ($M = 1.82$, $SD = 1.10$).

Discussion

In Experiment 3, we aimed at investigating if our ad-IAT can predict the choice and use of subtractive strategies when solving problems involving additive and subtractive solution options. In doing so, we were inspired by previous research showing that the IAT can successfully predict cognition, affect, and behavior (Greenwald et al., 2009).

Consistent with our hypothesis, we found a negative correlation between participants' IAT D-scores and the frequency with which they used subtracting transformations when working on the grids task (Adams et al., 2021). In other words, the more participants showed an (implicit) preference for additive concepts, the less likely they considered and employed "subtracting" problem solving strategies.

As mentioned in the Introduction, there is nothing inherently wrong with favoring additive approaches. The problem arises when this tendency leads individuals to overlook subtractive options, even when those options may be more efficient or otherwise preferable. This was precisely the case in Adams et al.'s study with grid task, where participants could not succeed (i.e., achieve the fewest mouse clicks possible) unless they actively thought about and employed subtractive transformations. Similar observations have been documented in relation to other cognitive biases. For instance, in the field of design, scholars have extensively explored the phenomenon of "design fixation," where individuals' adherence to specific design approaches hampers the exploration of potentially superior solutions (e.g. Jansson & Smith, 1991; Neroni, Vasconcelos, & Crilly, 2017). Likewise, research in the field of creativity demonstrated that prematurely committing to a specific process execution strategy can hinder the exploration of alternative (and potentially superior) solutions, ultimately undermining the generation of creative ideas (e.g. Neroni et al., 2017).

Just as Adams et al.'s speculated, people's tendency to favor "additive" actions may lead them to use an additive strategy more often. If this approach is perceived to be successful then it will become a common default when generating ideas to solve problems (e.g. Anderson, 1982; Ginossar & Trope, 1987; Higgins, 1996; Wiley, 1998). Indeed, in their study, Adams et al. found that participants relied on additive solution approaches especially under two conditions: (1) when high cognitive demands precluded the flexible pursuit of more tailored approaches, and (2) when there was an absence of information that cued alternative strategies (Adams et al., 2021; Fallen et al., 2022). The results of our Experiment 3 provide further support for Adams et al.'s hypothesis by showing that the level of additive bias (as revealed by our IAT) predicts how much people would be likely to "think and act in additive terms" and neglect subtractive options when solving problems.

Particularly relevant for interpreting our findings, previous research also showed that the IAT has the ability to predict cognitive processes at a lower level, specifically those related to automatic, bottom-up processes (Hugenberg & Bodenhausen, 2003). These processes indicate a heightened ability or tendency to process stimuli that align with one's implicit preferences. In the context of our experiment, this suggests that participants with a stronger additive bias, as indicated by their performance on the IAT, exhibited a higher threshold for recognizing alternative subtractive transformations while working on the grids task. In other words, they were less inclined to consider subtractive solutions as viable options. This hypothesis is consistent with the observations made by Adams et al. (2021) that the limited use of subtractive solutions by participants was not due to a lack of recognition of their value, but rather a failure to actively consider them (see Bilalić, M., McLeod, & Gobet, 2007 for similar results in a different context). Overall, these findings and related observations highlight the potential impact of the additive bias on decision-making processes and problem-solving.

Our results also replicated Adams et al.'s (2021) findings by showing that participants were more likely to produce a superior subtractive transformation when they had more opportunities to recognize the task-specific shortcomings of an additive search strategy. Indeed, we observed that more participants used subtractive transformations when solving the last "main grid" than when solving the first three "practice grids". However, the percentage of participants that used subtracting transformations in our experiment (i.e. 10% for the "practice grids" and 18% for the "main grid") was much lower than the percentage of participants that used subtracting transformations in Adams et al.'s study (Experiment 5: 48% for the "practice grids" and 63% for the "main grid"). We can speculate that this may be due to differences in sample size, but we are unable to provide here a definitive explanation for this observed difference.

Finally, from a methodological perspective, the results of Experiment 3 further validated the effectiveness and reliability of our ad-IAT as a research tool for investigating the additive bias by demonstrating its predictive capability to anticipate future choices and the corresponding behaviors.

General discussion

In the present study, drawing inspiration from the principles and methodologies of cognitive psychology, we aimed at developing a new digital tool, namely the additive bias IAT (ad-IAT), that can be used to investigate the additive bias.

In Experiment 1, we developed and validated the tool with a large sample of participants, from different cultures and linguistic backgrounds. Building upon this foundation, Experiment 2 entailed the creation and evaluation of an expanded version of the tool, the mad-IAT, designed to uncover multiple implicit mental associations underlying the additive bias. Finally, in Experiment 3, we further validated the efficacy of our tool by demonstrating that our ad-IAT has predictive value in anticipating future choices and behaviors. Overall, the results from our Experiments have theoretical and methodological implications.

From a theoretical standpoint, the findings presented here offer empirical evidence supporting the assumption that individuals exhibit a preference for addition over subtraction. Specifically, our results revealed that participants displayed an inherent tendency to assign a positive valence to additive concepts (Experiment 1 + extension) and to perceive additive actions as functional and safer (Experiment 2). In contrast, subtractive concepts were associated with a negative value (Experiment 1 + extension) and regarded as non-functional and risky (Experiment 2). This may seem perhaps not surprising, but the relevance of these findings is clear if one considers their implications. According to the theory of planned behavior (TPB, Ajzen, 1991) the most proximal predictor of behavior is the intention to perform it and one of the main drivers of intentions are attitudes, referring to the degree to which a person has a favorable or unfavorable evaluation of a given behavior. This may mean that if people have positive attitudes towards addition (and negative attitudes towards subtraction) they would be more inclined to perform additive choices and actions than their subtractive counterparts. This is exactly what we observed in our Experiment 3 where we found that implicit preference for addition resulted in participants favoring additive actions while neglecting subtractive alternatives when engaged in a problem-solving task.

What are the reasons behind people's preference for, and more frequent engagement in, additive actions compared to subtractive actions?

As highlighted by Adams et al. and discussed in this paper, the additive bias may be influenced by cultural, socioecological, and cognitive factors. From a socio-cultural perspective, over time, additive changes may be perceived more positively than subtractive changes (Adams et al., 2021; Glucksberg & McGlone, 2001; McAdams & de St. Aubin, 1992). Also, people may operate in an environment that may, probabilistically, offer more good opportunities to add than to subtract: originals require building before honing; the number of components that can be subtracted is always bound by what exists, and in designed environments, one may infrequently encounter artefacts from which the designers have not already subtracted the obviously negative components (Adams et al., 2021; Jannace & Camuffo, 2023).

From a cognitive perspective, people might be reluctant to subtract because of attentional and evaluative processes that favor the consideration and use of additive actions. Previous research showed that people routinely deviate from the 'rational choice' model of human behavior, in which one objectively weighs up the costs and benefits of all alternatives before choosing the optimal course of action (Gigerenzer & Gaissmaier, 2011). For example, people are often incapable of 'optimizing' (i.e., systematically processing all available information to maximize utility) and tend to choose options that are not necessarily the best solution to a problem, but rather the options that satisfies the requirements without extensive searching or consideration (Kahneman, 2003). This tendency, known as "satisficing" (Brown, 2004), facilitates more rapid, less effortful information processing, problem-solving and decision-making, while also affording potential wellbeing benefits (see Schwartz, Ward, Monterosso, Lyubomirsky, White, & Lehman, 2002). From this perspective, the additive bias may result from the higher availability of additive options compared to subtractive options. As Adams et al. (2021) pointed out, the general tendency to think and act in "additive" terms may lead people to use an additive strategy more often and with more perceived success making it a common default on which to rely when generating ideas to solve problems (e.g. Anderson, 1982; Ginossar & Trope, 1987; Higgins, 1996; Wiley, 1998).

Another mental heuristic that has been linked to the additive bias is the "status quo bias", defined as the tendency for "doing nothing or maintaining one's current or previous decision" (Samuelson & Zeckhauser, 1988). The additive bias, in this context, can be seen as a manifestation of the *status quo* bias, where individuals favor maintaining or adding to the existing state rather than subtracting or changing it. Regret avoidance and loss aversion are the most common explanations for *status quo* bias (Anderson, 2003; Kahneman, Knetsch, & Thaler, 1991). Regret avoidance suggests that individuals are motivated to avoid the regret they may experience if their decision turns out to be unfavorable (Brehm, 2007). Loss aversion, on the other hand, posits that people tend to place greater emphasis on avoiding losses rather than acquiring gains (Kahneman & Tversky, 1984; Tversky & Kahneman, 1991). As a result, they may be more inclined to maintain the current state to avoid potential losses associated with making changes.

Finally, the additive bias can be further exacerbated by the belief that subtractive choices and actions are less likely to be valued by others (Meyvis & Yoon, 2021). This aligns with perspectives in creativity research where, when introducing novelty to products or processes, proposing the elimination of something is often perceived as less creative compared to generating new additions (Fillon et al., 2022). Consistent with this view, studies within the framework of social inhibition indicate that situational variables, such as social loafing, conformity pressure and evaluation pressures, significantly contribute to inhibiting the consideration of "different" or "unpopular" courses of action when making decisions or generating ideas to solve a problem (Amabile, 1988; Amabile & Gryskiewicz, 1989; Magyari-Beck, 1992).

Additionally, individuals may assume that existing features or items serve a purpose, leading them to prioritize seeking additions as a more effective strategy. Furthermore, the sunk-cost bias (a tendency to persist in an endeavor after investing resources such as money, effort, or time) and aversion to waste may deter people from considering subtractive options, especially when they recognize or assume that the creation of those features or items required considerable effort to be created in the first place (e.g. Neroni & Crilly, 2019; Roth, Robbert, & Straus, 2015).

Overall, these perceived disadvantages of using subtractive strategies might encourage people to routinely seek out additive ones. Yet this tendency to "think in additive terms" may come at a price, with people often making worse decisions and poorer choices (or avoiding action altogether) when faced with more information or more options (Shafir & LeBoeuf, 2002; Tversky & Sharfir, 1992). As Adams et al.'s (2021) said "overlooking subtraction may mean that people are missing out on opportunities to make their lives more fulfilling, their institutions more effective and their planet more livable".

Historical and anecdotal evidence provide compelling examples that underscore the importance of considering subtractive options. We illustrated this in the Introduction with reference to McFarland's Strider bike and Kondo's observations about tidying up.

From a methodological perspective, our ad-IAT offers several advantages. Firstly, its flexibility allows for easy utilization in different languages and in different research settings, enabling the collection of normative data on the bias. For instance, the task can be published online and used to collect data from many online users, providing (1) research data on how demographic variables are associated with the magnitude of the bias, and (2) population data that can be compared with individual results using percentile scores for performance measures.

Secondly, owing to its ability to provide immediate and quantifiable performance metrics (with the task taking approximately 15 minutes), our ad-IAT can be readily employed to investigate the relationship between bias levels and individuals' behavior in various decision-making and problem-solving processes.

Lastly, our ad-IAT holds potential as an educational tool to raise awareness about the existence of the bias and motivate individuals to counteract it. The examples mentioned earlier, such as McFarland's invention and historical and anecdotal instances, clearly illustrate how the additive bias can restrict creativity and hinder idea generation in problem-solving, ultimately undermining innovation. Therefore, it is crucial to make people aware of this bias, as it is believed to occur automatically, particularly under conditions of high cognitive load (Adams et al., 2021), and to teach them how to guard against it. Previous research has shown that merely informing individuals about the presence of a bias fails to effectively inspire them to take action against it. One reason for this is that people often overlook, ignore, or underestimate their own personal susceptibility to risk or bias, compared to the vulnerability recognized within their respective group. This cognitive tendency is commonly referred to as the "bias blind spot" or "bias bias" (Pronin, Lin, & Ross, 2002; West, Meserve, & Stanovich, 2012). Furthermore, there is often a discrepancy between what individuals know and what they actually do in many domains of human behavior, resulting in knowledge-action gaps (e.g., Sligo & Jameson, 2000; Kennedy, Regehr, Rosenfield, & Lingard, 2004), value-action gaps (e.g., Blake, 1999), attitude-action gaps (Boulstridge & Carrigan, 2000), and/or intention-action gaps (Sheeran & Webb, 2016; Sheeran, 2011; Kollmuss & Agyeman, 2002). A relevant domain where this discrepancy is evident is residential energy use (Kennedy, Beckley, McFarlane, & Nadeau, 2009; Kollmuss & Agyeman, 2002; Abrahamse, & Steg, 2011). Many individuals express concern about climate change and understand the importance of conserving energy, yet their knowledge and attitudes do not translate into efforts to reduce energy consumption (Frederiks et al., 2015). Thus, the question arises: how can we effectively educate people about the existence of the additive bias and motivate them to consider subtractive alternatives?

Previous studies suggest that an effective approach to raising awareness of a bias and inspiring action against it is to demonstrate individuals' personal susceptibility to the bias itself (e.g., Neroni & Crilly, 2020; Norris, Smith, & Kaniasty, 1999; Sagarin, Cialdini, Rice, & Serna, 2002). This approach has been employed for years by utilizing the IAT to raise individuals' awareness of their own biases as a means of guarding against them in the future (Devine, Forscher, Austin, & Cox, 2012). This observation underscores the potential of our ad-IAT as a means of raising people's awareness of their own bias levels. Following this "raising awareness phase", it would be beneficial to provide concrete examples that highlight how this bias can impede idea generation and hinder problem-solving abilities, ultimately resulting in suboptimal choices. Prior work suggests that this would reduce the prevalence of the bias or its magnitude.

Limitations

We acknowledge the limitations of our experiments, particularly from a methodological perspective.

Firstly, the IAT has often been criticized for reflecting cultural and societal expectations rather than personal beliefs (Olson & Fazio, 2004). However, it is important to note that, whatever the reason underlying the observed attitudes, they can still impact behavior (Banaji, Nosek, & Greenwald, 2004). The results from our Experiment 3 further confirm this point by showing that individuals' preferences for additive concepts translated into more additive actions when solving problems.

Secondly, we developed and employed a word-based version of the IAT, which may have limitations associated with the use of linguistic terms (e.g., familiarity with the words used, level of literacy, verbal fluency, reading speed, etc.) and their mental representations. However, despite these potential constraints, the consistent IAT effects observed in Experiment 1 (and its extension) across diverse cultures and linguistic backgrounds corroborate the reliability of this verbal version as a means of examining the additive bias. In forthcoming investigations, it would be valuable to explore alternative versions of the proposed ad-IAT, such as incorporating different word stimuli or incorporating non-verbal stimuli, in order to validate their consistency with our findings.

Thirdly, it is important to acknowledge that our experimental design does not entirely eliminate or provide specific insights into the influence of image schemas on participants' performance. Prior research has indeed highlighted that the general population associates addition with right/up directions, and subtraction with left/down directions (e.g., Richardson et al., 2001). To minimize the potential confounding effects of these image schemas on participants' performance, we implemented a comprehensive counterbalancing procedure in our IATs: one group of participants completed the task with inverted target labels (from Block 5 to Block 7), while the other group completed the task with inverted attribute labels (from Block 5 to Block 7). Despite employing this counterbalancing approach, we were unable to cover all possible combinations among target stimuli, attribute stimuli, and spatial positions. For instance, none of our trials required participants to use the right key assignment to categorize the combination of "add" stimuli and "pleasant" stimuli. To thoroughly investigate the interplay between additive/subtractive concepts and spatial positions, future studies could implement full counterbalancing of the target and attribute labels. This might have important implications for how IATs and other similar experimental paradigms are designed and implemented, so that key position does not introduce a confounding variable.

Another set of limitations stems from the data collection approach employed in our study. Specifically, all the experiments conducted in this research were administered online. While the Covid-19 pandemic has led to an increase in online data collection in recent years, prompting researchers to reconsider the methodological advantages of online studies (e.g., Buhrmester, Talaifar, & Gosling, 2018; Carpenter et al., 2019), the issue of low-quality data remains a significant concern in online research (Gosling & Mason, 2015). However, it is worth noting that the scoring procedure for the IAT partially addresses this concern. For instance, the D-score drop mechanism (Nosek & Banaji, 2001) is designed to identify participants who rapidly and haphazardly press buttons to bypass the IAT. This mechanism helps in flagging such participants, thereby mitigating the impact of low-quality data to some extent.

In addition, to measure reaction times, we followed the approach outlined by Matthijsse, De Leeuw, and Hox (2015), recording timestamps at the start and end of each trial and calculating the differences in milliseconds. This procedure has been validated by Reimers and Stewart (2015). They found that across various machines, with reaction times of approximately 300–600 ms, latencies were consistently similar within each machine (within-system SDs ≈ 10 ms) and were inflated by approximately 50–80 ms. Differences across web browsers were negligible (typically < 10 ms). Therefore, they concluded that although a given latency might be slightly inflated, the degree of inflation should be relatively constant for a given machine (and therefore participants), rendering it largely irrelevant for within-subjects tests like the IAT.

Fourthly, while we are encouraged by the results from Experiment 3, which demonstrate that our ad-IAT can predict future choices and behaviors, we acknowledge that this observation was limited to a problem-solving task with specific characteristics. We chose this task because of its adaptability for online administration and because, to our knowledge, it is the only available and validated task that can empirically test preferences for additive choices over subtraction choices (Adams et al., 2021). However, we acknowledge that the grids task employed in Experiment 3 is relatively abstract and may not accurately reflect real-world problem-solving scenarios, thus lacking ecological validity (Finn et al., 2022). This raises concerns regarding the extent to which these findings can be generalized to real-life behaviors. Future studies could more deeply investigate the relationship between the additive bias and various types of decision-making and problem-solving processes.

Finally, one could contend that the reliability of our findings regarding the magnitude of the additive bias may stem from an experimental artifact. Previous research has indicated that implicit attitudes, as measured by the IAT, can be "unstable," implying that they are susceptible to change through experimental manipulations (e.g., Fiedler, Messner, & Bluemke, 2006). However, the consistency of our results across multiple experiments, despite procedural modifications, challenges this notion. Moreover, the crucial revelation that the bias level, as measured by our task, effectively predicts individuals' behaviors in problem-solving supports the validity of the claims we make about our findings. This suggests that the additive bias is not merely a transient outcome of experimental conditions but rather a robust and influential factor impacting individuals' cognitive processes in the context of problem-solving.

Conclusions

By developing and employing an additive bias IAT, we shed additional light on the reasons behind people's preference for additive actions. Our series of experiments demonstrated the efficacy of the IAT in uncovering and quantifying this bias, providing a deeper understanding of its underlying factors and its influence on people's behavior.

The implications of our findings extend to various domains such as decision-making, problem-solving, creativity, and, more generally, innovation. Specifically, the additive bias can constrain individuals' decision-making and problem-solving abilities by prioritizing the addition of new elements over the removal or simplification of existing ones. This inclination hampers the exploration of more efficient solutions. In fields like design, this bias can lead to the creation of unnecessarily complex or intricate systems. Incorporating features without considering overall complexity can result in inefficiencies, increased costs, and challenges in maintenance and scalability.

The preference for additive actions can engender resistance to change. Innovations that involve the removal or replacement of established elements may face heightened resistance due to the perception that subtractive choices are undervalued or less likely to be appreciated. Furthermore, the additive bias can have environmental consequences leading people to accumulate resources and features, potentially leading to wastefulness. This accumulation mindset may hinder the exploration of subtractive alternatives that could lead to more sustainable and resource-efficient outcomes.

To address the additive bias and foster a wider array of effective and optimal approaches to innovation, it is therefore vital to develop methodologies that explicitly identify and counteract the causes and consequences of the additive bias. By acknowledging and mitigating the impact of this bias, we can encourage thinking that embraces both addition and subtraction, leading to outcomes that are more efficient and sustainable. Our research lays the foundation for future investigations into the additive bias and for the development of strategies aimed at mitigating its impact.

References

Abrahamse, W., & Steg, L. (2011). Factors related to household energy use and intention to reduce it: The role of psychological and socio-demographic variables. *Human ecology review*, 30-40.

Adams, G. S., Converse, B. A., Hales, A. H., & Klotz, L. E. (2021). People systematically overlook subtractive changes. *Nature*, *592*(7853), 258-261. https://doi.org/10.1038/s41586-021-03380-y.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T.

Amabile, T. M. (1988). A model of creativity and innovation in organizations. *Research in organizational behavior*, *10*(1), 123-167.

Amabile, T. M., & Gryskiewicz, N. D. (1989). The creative environment scales: Work environment inventory. *Creativity research journal*, *2*(4), 231-253. https://doi.org/10.1080/10400418909534321.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, *89*(4), 369. https://doi.org/10.1037/0033-295X.89.4.369.

Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction. *The International Review of Research in Open and Distributed Learning*, *4*(2). https://doi.org/10.19173/irrodl.v4i2.149.

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes and Tetlock. *Psychological inquiry*, *15*(4), 279-289.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für experimentelle Psychologie*, *48*(2), 145-160. https://doi.org/10.1026/0949-3946.48.2.145.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological science*, *17*(1), 53-58. 10.1111/j.1467-9280.2005.01664.x.

Bertamini, M., Makin, A., & Rampone, G. (2013). Implicit association of symmetry with positive valence, high arousal and simplicity. *i-Perception*, *4*(5), 317-327. https://doi.org/10.1068/i0601jw.

Bilalić, M., McLeod, P., & Gobet, F. (2007). Does chess need intelligence?—A study with young chess players. *Intelligence*, *35*(5), 457-470. https://doi.org/10.1016/j.intell.2006.09.005.

Bilalić, M., McLeod, P., & Gobet, F. (2010). The mechanism of the Einstellung (set) effect: A pervasive source of cognitive bias. *Current Directions in Psychological Science*, *19*(2), 111-115. https://doi.org/10.1177/0963721410363571.

Blake, J. (1999). Overcoming the 'value-action gap'in environmental policy: Tensions between national policy and local experience. *Local environment*, *4*(3), 257-278. https://doi.org/10.1080/13549839908725599.

Boulstridge, E., & Carrigan, M. (2000). Do consumers really care about corporate responsibility? Highlighting the attitude—behaviour gap. *Journal of communication management*, *4*(4), 355-368. https://doi.org/10.1108/eb023532.

Braess, D., Nagurney, A., & Wakolbinger, T. (2005). On a paradox of traffic planning. *Transportation science*, *39*(4), 446-450. https://doi.org/10.1287/trsc.1050.0127.

Brandenstein, N., Gebauer, F., & Carbon, C. C. (2019). How Do We Perceive “Aliens”? About the Implicit Processes Underlying the Perception of People With Alien Paraphernalia. *Frontiers in psychology,* 10, 1551. https://doi.org/10.3389/fpsyg.2019.01551.

Brehm, J. W. (2007). A brief history of dissonance theory. *Social and Personality Psychology Compass*, *1*(1), 381-391. https://doi.org/10.1111/j.1751-9004.2007.00035.x.

Brown, R. (2004). Consideration of the origin of Herbert Simon's theory of “satisficing”(1933-1947). *Management Decision*, *42*(10), 1240-1256. https://doi.org/10.1108/00251740410568944.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon’s Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, *13*(2), 149-154. https://doi.org/10.1177/1745691617706516.

Carpenter, T. P., Pogacar, R., Pullig, C., Kouril, M., Aguilar, S., LaBouff, J., Isenberg, N., & Chakroff, A. (2019). Survey-software implicit association tests: A methodological and empirical analysis. *Behavior research methods*, *51*, 2194-2208. https://doi.org 10.3758/s13428-019-01293-3.

Collino, S., & Lauto, G. (2022). Reducing cognitive biases through digitally enabled training. A conceptual framework. In *Do Machines Dream of Electric Workers? Understanding the Impact of Digital Technologies on Organizations and Innovation* (pp. 179-191). Springer International Publishing. https://doi.org/10.1007/978-3-030-83321-3_13.

Crilly, N., & Cardoso, C. (2017). Where next for research on fixation, inspiration and creativity in design?. *Design Studies*, *50*, 1-38. https://doi.org/10.1016/j.destud.2017.02.001.

Cristofaro, M., Leoni, L., & Giardino, P. L. (2021). Cognitive biases' influence on employees' product creativity and product performance: evidences from Italian manufacturing technology firms. *Journal of Manufacturing Technology Management*, *33*(4), 675-695. https://doi.org/10.1108/JMTM-10-2020-0414.

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin*, *29*(4), 532-544. https://doi.org/10.1177/0146167202250923.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, *48*(6), 1267-1278. https://doi.org/10.1016/j.jesp.2012.06.003.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an Implicit Association Test for assessing anxiety. *Journal of personality and social psychology*, *83*(6), 1441. https://doi.org/10.1037/0022-3514.83.6.1441.

Eidelman, S., & Crandall, C. S. (2009). A psychological advantage for the status quo. *Social and psychological bases of ideology and system justification*, 85-106. https://doi.org/10.1093/acprof:oso/9780195320916.003.004.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and brain sciences*, *32*(5), 429-448. https://doi.org 10.1017/S0140525X0999094X.

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European review of social psychology*, *17*(1), 74-147. https://doi.org/10.1080/10463280600681248.

Fillon, A., Girandola, F., Bonnardel, N., & Souchet, L. (2022). People systematically overlook subtractive changes (2021): Replication and extension. https://doi.org/10.31234/osf.io/4jkvn.

Fischer, M. H., Winter, B., Felisatti, A., Myachykov, A., Mende, M. A., & Shaki, S. (2021). More instructions make fewer subtractions. *Frontiers in Psychology*, *12*, 720616. https://doi.org/10.3389/fpsyg.2021.720616.

Frederiks, E. R., Stenner, K., & Hobman, E. V. (2015). Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour. *Renewable and Sustainable Energy Reviews*, *41*, 1385-1394. https://doi.org/10.1016/j.rser.2014.09.026.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The journal of socio-economics*, *40*(1), 35-42. https://doi.org/10.1016/j.socec.2010.10.008.

Gallo, C. (2011). *Innovation Secrets of Steve Jobs: Insanely Different Principles for Breakthrough Success*. McGraw-Hill Education.

Gattol, V., Sääksjärvi, M., & Carbon, C. C. (2011). Extending the implicit association test (IAT): assessing consumer attitudes based on multi-dimensional implicit associations. *PLoS one*, *6*(1), e15849. https://doi.org/10.1371/journal.pone.0015849.

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology, 49*(3), 171–180. https://doi.org/10.1026/1618-3169.49.3.171.

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology*, *33*(5), 573-589. https://doi.org/10.1002/ejsp.166.

Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of abnormal psychology*, *110*(2), 282. https://doi.org/10.1037//0021-843x.110.2.282.

Gibbs Jr, R. W. (1996). Why many concepts are metaphorical. *Cognition*, *61*(3), 309-319. https://doi.org/10.1016/s0010-0277(96)00723-8.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, *62*, 451-482. https://doi.org/10.1146/annurev-psych-120709-145346.

Gilbert, D. T., & Ebert, J. E. (2002). Decisions and revisions: the affective forecasting of changeable outcomes. *Journal of personality and social psychology*, *82*(4), 503. 10.1037//0022-3514.82.4.503.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and social Psychology*, *52*(3), 464. https://doi.org/10.1037/0022-3514.52.3.464.

Glucksberg, S., & McGlone, M. S. (2001). *Understanding figurative language: From metaphor to idioms* (No. 36). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195111095.001.0001.

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual review of psychology*, *66*, 877-902. https://doi.org/10.1146/annurev-psych-010814-015321.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*(6), 1464. https://doi.org/10.1037/0022-3514.74.6.1464.

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48*(2), 85–93. https://doi.org/10.1026/0949-3946.48.2.85.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of personality and social psychology*, *85*(2), 197. https://doi.org/10.1037/0022-3514.85.2.197.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology*, *97*(1), 17. https://doi.org/10.1037/a0015575.

Higgins, E. T. (1996). *Social Psychology: Handbook of Basic Principles* (eds Higgins, E. T. & Kruglanski, A. W.). 75–123. Guilford Publications.

Highhouse, S., & Zhang, D. (2015). The new fruit fly for applied psychological research. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8*(2), 179–183. https://doi.org/10.1017/iop.2015.22.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, *14*(6), 640-643. https://doi.org/10.1046/j.0956-7976.2003.psci_1478.x.

Jannace, D., & Camuffo, A. (2023). Curbing the Addition Bias: Scientific Approaches and Propensity to Subtract in Entrepreneurial Ideas. In *Academy of Management Proceedings* (Vol. 2023, No. 1, p. 15426). Briarcliff Manor, NY 10510: Academy of Management. https://doi.org/10.5465/AMPROC.2023.194bp.

Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design studies*, *12*(1), 3-11. https://doi.org/10.1016/0142-694X(91)90003-F

Johnson, J. (2022). Delegating strategic decision-making to machines: Dr. Strangelove Redux?. *Journal of Strategic Studies*, *45*(3), 439-477. https://doi.org/10.1080/01402390.2020.1759038.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697. https://doi.org/10.1037/0003-066X.58.9.697.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of political Economy*, *98*(6), 1325-1348. https://doi.org/10.1086/261737.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, *5*(1), 193-206. https://doi.org/10.1257/jep.5.1.193.

Kahneman, D., & Tversky, A. (1979). On the interpretation of intuitive probability: A reply to Jonathan Cohen. https://doi.org/10.1016/0010-0277(79)90024-6.

Kennedy, E. H., Beckley, T. M., McFarlane, B. L., & Nadeau, S. (2009). Why we don't" walk the talk": Understanding the environmental values/behaviour gap in Canada. *Human Ecology Review*, 151-160. https://www.jstor.org/stable/24707539.

Kennedy, T., Regehr, G., Rosenfield, J., Roberts, S. W., & Lingard, L. (2004). Exploring the gap between knowledge and behavior: a qualitative study of clinician action following an educational intervention. *Academic Medicine*, *79*(5), 386-393. https://doi.org/10.1097/00001888-200405000-00006.

Klotz, L. (2021). *Subtract: The untapped science of less*. Flatiron Books.

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior?. *Environmental education research*, *8*(3), 239-260. https://doi.org/10.1080/13504620220145401.

Kondo, M. (2014). *The life-changing magic of tidying: A simple, effective way to banish clutter forever*. Random House.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV. *Implicit measures of attitudes*, *59*, 102. https://doi.org/10.1037/0022-3514.85.2.197.

Liedtka, J. (2015). Perspective: Linking design thinking with innovation outcomes through cognitive bias reduction. *Journal of product innovation management*, *32*(6), 925-938. https://doi.org/10.1111/jpim.12163.

Lovallo, D., & Kahneman, D. (2000). Living with uncertainty: Attractiveness and resolution timing. *Journal of Behavioral Decision Making*, *13*(2), 179-190. https://doi.org/10.1002/(SICI)1099-0771(200004/06)13:2<179::AID-BDM332>3.0.CO;2-J.

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in cognitive sciences*, *24*(11), 930-944. https://doi.org/10.1016/j.tics.2020.08.005.

Magyari-Beck, I. (1992). Identifying the blocks to creativity in Hungarian culture. *Creativity Research Journal*, *5*(4), 419-427. https://doi.org/10.1080/10400419209534455.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 32*(1), 61–69.

Majid, A., Bowerman, M., Kita, S., Haun, D. B., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in cognitive sciences*, *8*(3), 108-114. https://doi.org/10.1016/j.tics.2004.01.003

Matthijsse, S. M., De Leeuw, E. D., & Hox, J. J. (2015). Internet panels, professional respondents, and data quality. *Methodology*. https://doi.org/10.1027/1614-2241/a000094.

McAdams, D. P., & de St Aubin, E. D. (1992). A theory of generativity and its assessment through self-report, behavioral acts, and narrative themes in autobiography. *Journal of personality and social psychology*, *62*(6), 1003. https://doi.org/10.1037/0022-3514.62.6.1003.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of experimental Social psychology*, *37*(5), 435-442. https://doi.org/10.1006/jesp.2000.1470.

Meyvis, Tom, and Heeyoung Yoon. "Adding is favoured over subtracting in problem solving." (2021): 189-190. https://doi.org/10.1038/d41586-021-00592-0.

Mueller, J. S., Melwani, S., & Goncalo, J. A. (2012). The bias against creativity: Why people desire but reject creative ideas. *Psychological science*, *23*(1), 13-17. https://doi.org/10.1177/0956797611421018.

Mumford, M. D., Blair, C., Dailey, L., Leritz, L. E., & Osburn, H. K. (2006). Errors in creative thought? Cognitive biases in a complex processing activity. *The Journal of Creative Behavior*, *40*(2), 75-109. https://doi.org/10.1002/j.2162-6057.2006.tb01267.x.

Neroni, M. A., & Crilly, N. (2019). Whose ideas are most fixating, your own or other people's? The effect of idea agency on subsequent design behaviour. *Design Studies*, *60*, 180-212. https://doi.org/10.1016/j.destud.2018.05.004.

Neroni, M. A., & Crilly, N. (2020). Enhancing creativity by demonstrating individual vulnerability to fixation. In *Proceedings of the Sixth International Conference on Design Creativity (ICDC 2020)* (pp. 060-067). https://doi.org/10.17863/CAM.51149.

Neroni, M. A., Vasconcelos, L. A., & Crilly, N. (2017). Computer-based “mental set” tasks: An alternative approach to studying design fixation. *Journal of Mechanical Design*, *139*(7), 071102. https://doi.org/10.1115/1.4036562.

Neumann, R., Hülsenbeck, K., & Seibt, B. (2004). Attitudes towards people with AIDS and avoidance behavior: Automatic and reflective bases of behavior. *Journal of Experimental Social Psychology*, *40*(4), 543-550. https://doi.org/10.1016/j.jesp.2003.10.006.

Norris, F. H., Smith, T., & Kaniasty, K. (1999). Revisiting the experience–behavior hypothesis: The effects of Hurricane Hugo on hazard preparedness and other self-protective acts. *Basic and Applied Social Psychology*, *21*(1), 37-47.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social cognition*, *19*(6), 625-666. https://doi.org/10.1521/soco.19.6.625.20886.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. *Automatic processes in social thinking and behavior*, *4*, 265-92.

Olson, M. A., & Fazio, R. H. (2004). Trait inferences as a function of automatically activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology*, *26*(1), 1-11. https://doi.org/10.1207/s15324834basp2601_1.

Oswald, M. E., & Grosjean, S. (2004). Confirmation bias. *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*, *79*. https://doi.org/10.13140/2.1.2068.0641.

Palfai, T. P., & Ostafin, B. D. (2003). Alcohol-related motivational tendencies in hazardous drinkers: Assessing implicit response tendencies using the modified-IAT. *Behaviour research and therapy*, *41*(10), 1149-1162. https://doi.org/10.1016/S0005-7967(03)00018-4.

Page, T. (2009). Feature creep and usability in consumer electronic product design. *International Journal of Product Development*, *9*(4), 406-428. https://doi.org/10.1504/IJPD.2009.027474.

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, *28*(3), 369-381. https://doi.org/10.1177/0146167202286008.

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior research methods*, *47*, 309-327. https://doi.org/10.3758/s13428-014-0471-1.

Richardson, D. C., Spivey, M. J., Edelman, S., & Naples, A. J. (2001). " Language is spatial": Experimental evidence for image schemas of concrete and abstract verbs. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 23, No. 23). https://doi.org/10.1037/xlm0000149.

Roth, S., Robbert, T., & Straus, L. (2015). On the sunk-cost effect in economic decision-making: a meta-analytic review. *Business research*, *8*, 99-138. https://doi.org/10.1007/s40685-014-0014-8.

Rudman, L. A., Feinberg, J., & Fairchild, K. (2002). Minority members' implicit attitudes: Automatic ingroup bias as a function of group status. *Social Cognition*, *20*(4), 294-320. https://doi.org/10.1521/soco.20.4.294.19908.

Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: the motivations and mechanisms of resistance to persuasion. *Journal of personality and social psychology*, *83*(3), 526. https://doi.org/10.1037/0022-3514.83.3.526.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, *1*, 7-59. https://www.jstor.org/stable/41760530.

Schnabel, K., Asendorpf, J. B., & Greenwald, A. G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment*, *24*(4), 210-217. https://doi.org/10.1027/1015-5759.24.4.210.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: happiness is a matter of choice. *Journal of personality and social psychology*, *83*(5), 1178. https://doi.org/10.1037/0022-3514.83.5.1178.

Schultz, P. W., Shriver, C., Tabanico, J. J., & Khazian, A. M. (2004). Implicit connections with nature. *Journal of Environmental Psychology, 24*(1), 31–42. https://doi.org/10.1016/S0272-4944(03)00022-7.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual review of psychology*, *53*(1), 491-517. https://doi.org/10.1146/annurev.psych.53.100901.135213.

Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. *Social and personality psychology compass*, *10*(9), 503-518. https://doi.org/10.1111/spc3.12265.

Sligo, F. X., & Jameson, A. M. (2000). The knowledge—behavior gap in use of health information. *Journal of the American Society for Information Science*, *51*(9), 858-869. https://doi.org/10.1002/(SICI)1097-4571(2000)51:9<858::AID-ASI80>3.0.CO;2-Q.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory* (Vol. 1). Princeton university press. https://doi.org/10.1046/j.1420-9101.1988.1010086.x.

Thompson, D. V., Hamilton, R. W., & Rust, R. T. (2005). Feature fatigue: When product capabilities become too much of a good thing. *Journal of marketing research*, *42*(4), 431-442. https://doi.org/10.1509/jmkr.2005.42.4.431.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The quarterly journal of economics*, *106*(4), 1039-1061. https://doi.org/10.2307/2937956.

Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological science*, *3*(6), 358-361. https://doi.org/10.1111/j.1467-9280.1992.tb00047.x.

Veblen, T. (1994). *The Theory of the Leisure Class: An Economic Study of Institutions* . Dover.

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of personality and social psychology*, *103*(3), 506. https://doi.org/10.1037/a0028857.

Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition, 26*(4), 716–730. https://doi.org/10.3758/BF03211392.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253-265. https://doi.org/10.1002/wcs.104.