Computers in Human Behavior: Artificial Humans 8 (2026) 100290

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

Research paper

Technology paternalism and the reactance deficit — Are our natural protective mechanisms failing us?

Michelle Görlitz[1], Astrid Marieke Rosenthal-von der Pütten[*,1]

iTec – Chair Individual and Technology, RWTH Aachen University, Germany

ARTICLE INFO

Dataset link: RawDataPaternalismStudy2024Anonymized.xlsx (Original data)

Keywords:
Technology paternalism
Human-AI interaction
Reactance
Autonomy

ABSTRACT

With the rapid advancement of artificial intelligence (AI), the potential for paternalistic system behavior increases. While restrictions on autonomy are known to undermine psychological well-being, AI systems introduce a particular challenge: their opacity may prevent users from recognizing when and how their autonomy is constrained. As a result, autonomy restrictions may go unnoticed, potentially weakening natural protective mechanisms such as psychological reactance. To investigate the relationship between AI paternalism, autonomy perceptions, and reactance, we conducted a mixed-design online experiment (N = 137) in an e-commerce context, manipulating five levels of AI influence (control, nudging, reversible restriction, non-reversible restriction, covert restriction). Autonomy perceptions and reactance were assessed before and after a belief-updating intervention disclosing the system's paternalistic mechanisms. Results show that covert restrictions initially elicited comparatively weak reactance, whereas salient nudges triggered elevated reactance despite preserving choice. Belief updating partially recalibrated autonomy perceptions, particularly when prior restrictions had remained opaque. Mediation analyses support a dual-pathway model of technology reactance: changes in perceived personal autonomy constitute the primary mechanism underlying reactance, while increased attribution of system autonomy provides an additional pathway. These findings suggest that reactance in AI contexts depends less on restriction per se than on how autonomy threats are cognitively represented. We propose a dual-pathway framework linking paternalism perception, belief updating, and reactance to guide future research on autonomy-preserving AI design.

AI systems are increasingly shaping human decisions by filtering options, nudging behavior, or preventing actions altogether. Navigation systems silently reroute drivers, recommender systems restrict exposure to alternatives, and clinical decision-support systems prevent physicians from selecting treatments flagged as high risk. While these systems are often designed to improve outcomes, such interventions potentially reduce users' autonomy and do so frequently without their awareness. These types of interventions can be conceptualized as forms of technology paternalism. Technology paternalism refers to situations in which a technology, acting on behalf of a patron who embeds moral or normative values into the system, interferes with a user's autonomy without explicit consent in order to promote the user's perceived welfare or best interests. The defining feature of technology paternalism, and the central problem addressed in this work, is that such autonomy restrictions often operate opaquely. In contrast, traditional forms of paternalism exerted by parents, medical professionals, or governments are typically explicit and attributable to identifiable actors. Parents prohibit risky sports and parties and limit their children's screen time. Physicians withhold or structure information in the interest of patient welfare. Governments protect their citizens by introducing laws that restrict their autonomy, such as mandatory use of security belts in cars and prohibition of drugs. Although these measures restrict autonomy, they are usually visible and socially recognizable as interventions. Psychological Reactance Theory (PRT; Brehm, 1966) posits that perceived threats to autonomy trigger motivational states aimed at restoring freedom. Crucially, reactance depends on the perception of restriction. If paternalistic interventions remain unnoticed the protective mechanism of reactance may not be activated. Consequently, autonomy effects in the context of technology paternalism may differ fundamentally from those observed in traditional paternalistic settings.

Although prior research has examined autonomy in human–AI interaction (Mahmud et al., 2022) as well as nudging (Thaler & Sunstein, 2021) and persuasive technologies (Oinas-Kukkonen & Harjumaa, 2008), it remains unclear under which conditions autonomy restrictions actually elicit psychological reactance. In particular, existing

* Correspondence to: Chair Individual and Technology, RWTH Aachen University, Theaterplatz 14, 52062 Aachen, Germany.
E-mail address: arvdp@itec.rwth-aachen.de (A.M. Rosenthal-von der Pütten).
[1] Equally contribution authors.

https://doi.org/10.1016/j.chbah.2026.100290
Received 18 July 2025; Received in revised form 27 February 2026; Accepted 16 March 2026
Available online 31 March 2026
2949-8821/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

work rarely disentangles the objective presence of paternalistic intervention from users' subjective perception of it, and the distinction between overt and covert forms of restriction remains underexplored. Addressing these gaps is essential for understanding how technology paternalism influences autonomy and user responses.

In this work, we address these gaps through three primary contributions. First, we refine the concept of technology paternalism by proposing a definition that distinguishes between the ontological presence of paternalistic intervention and its epistemic perception by users, explicitly modeling the socio-technical relationship between technological systems and their human patrons (developers). Second, we develop a theoretical framework linking technology paternalism, perceived autonomy, and psychological reactance, with emphasis on false-positive and false-negative classifications arising from opaque AI behavior. Third, we provide empirical evidence from a controlled experimental study comparing overt, covert, and non-restrictive AI systems, investigating how different levels of paternalistic influence affect autonomy perceptions and reactance, and how belief updating changes these effects. Together, these contributions advance understanding of when and how AI-mediated autonomy restrictions influence user responses and offer implications for the design of human-centered AI systems.

1. Related work

1.1. Traditional paternalism and reactance

Autonomy restrictions are a pervasive feature of social life, spanning private and public domains (Dworkin, 2020). Individuals encounter such restrictions early in their life with rules and limits set by parental figures and later with institutional regulations such as speed limits or restrictions on certain substances. Although these interventions limit individual freedom, they often follow a paternalistic logic: constraining autonomy in order to promote welfare or prevent harm. Accordingly, paternalism has been extensively discussed in moral philosophy as well as across psychology, management, marketing, and related disciplines (Rochi, 2023).

Given the diversity of definitions in the literature, it is important to establish a working understanding of paternalism. The American Psychological Association defines paternalism as "a policy or attitude in which those having authority over others extend this authority into areas usually left to individual choice or conscience (e.g., smoking, sexual behavior), usually on the grounds that this is necessary for the welfare or protection of the individuals concerned." (APA Dictionary of Psychology, 2018).

In a widely cited philosophical analysis, Dworkin (2020) proposes "the following conditions as an analysis of *X acts paternalistically toward Y by doing (omitting) Z*:

1. Z (or its omission) interferes with the liberty or autonomy of Y.
2. X does so without the consent of Y.
3. X does so only because X believes Z will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of Y".

For instance, lying to prevent someone from engaging in risky behavior satisfies the first condition by interfering with informed choice; withholding information or acting without consulting the individual illustrates the absence of consent; and mandatory safety regulations exemplify actions justified by the intention to protect welfare. Importantly, acting without consent does not only include explicit opposition but also situations in which individuals are unaware of the intervention or were not given the opportunity to consent. Thus, some paternalistic actions may restrict autonomy without being recognized as such by the affected individual.

From the perspective of Self-Determination Theory (SDT, Deci & Ryan, 1985), autonomy constitutes a basic psychological need essential for well-being (Kachanoff et al., 2019; Vansteenkiste et al., 2020a). Consequently, paternalistic interventions are not only ethical or normative issues but also psychological ones, because autonomy restrictions can influence perceived agency, motivation, and well-being. Understanding how individuals respond to autonomy threats therefore requires consideration of underlying psychological mechanisms.

The well-established PRT (Brehm, 1966) provides a central framework for understanding responses to perceived autonomy restrictions. The classic process of eliciting reactance reactions starts with a state of freedom, which is then threatened or eliminated. This in turn triggers the motivational state of reactance and according reactions, which can lead to freedom restoration attempts (Dillard & Shen, 2005). Elimination of freedom is blocking the person from performing an action, while "anything that impedes, but does not; eliminate, freedom is a threat (e.g., attempted social influence, Brehm (1966))" (Rosenberg & Siegel, 2018a, p.4). This simplified process feeds the assumption that intentional freedom restoration activity is necessarily preceded by the motivational state of reactance, making reactance an essential concept for an autonomous life. In this context, the term *freedom* is frequently used and refers to behavioral freedom, which we will use synonymously to *autonomy*, defined as "the experience of acting from choice" based on self-determination theory (SDT, Deci & Ryan, 1985). In summary, within this framework, reactance functions as a mechanism through which individuals defend their autonomy.

A central assumption of PRT is that reactance emerges only when a threat to freedom *is perceived*, but autonomy restrictions differ in their visibility and obviousness. While some interventions are explicit and transparent (e.g., parents communicating rules), others may be subtle or difficult to detect (e.g., physicians omitting information on certain treatments), potentially blurring the boundary between guidance and manipulation. Consequently, psychological responses to paternalistic interventions depend not solely on whether autonomy is objectively restricted but on whether such restriction is perceived as threatening.

While traditional paternalistic interventions are often explicit and attributable to identifiable actors, emerging AI systems introduce forms of paternalistic influence that may operate opaquely. This raises the question of how technological contexts alter the perception and psychological consequences of autonomy restriction, which we address in the following section.

1.2. Technology paternalism and reactance

1.2.1. Defining technology paternalism

While paternalism has traditionally been analyzed in interpersonal or institutional contexts, emerging digital and AI systems introduce novel forms of autonomy-restricting interventions. Research on persuasive technologies (Fogg, 1998; Oinas-Kukkonen & Harjumaa, 2008), nudging (Thaler & Sunstein, 2021), and algorithmic decision-making (Mahmud et al., 2022) has implicitly engaged with paternalistic dynamics, even when the term itself is avoided presumably due to a negative connotation and the implicated normative debate. Within this broader landscape, technology paternalism refers to situations in which a technology, acting on behalf of a patron who embeds moral or normative values into the system, interferes with a user's autonomy without explicit consent in order to promote the user's perceived welfare or best interests.

Persuasive technologies and nudging approaches are often described as instances of libertarian or soft paternalism, as they aim to steer behavior toward beneficial outcomes while still preserving freedom of choice (Hanna, 2018; Thaler & Sunstein, 2021). However, it has been debated that some types of nudges are indeed paternalistic due to their manipulative nature (Barton & Grüne-Yanoff, 2015). In the present work, nudging is considered a potential form of soft technology paternalism insofar as it attempts to influence decision-making for presumed welfare-related reasons. In the area of algorithmic decision-support-systems, the phenomenon of algorithm aversion describes the

tendency to disregard or override algorithmic recommendations in favor of human judgment. Existing models (e.g., Burton et al., 2020; Mahmud et al., 2022) explain algorithm aversion primarily through trust deficits, perceived competence, and contextual appropriateness. While Mahmud et al. (2022) include psychological factors in their framework and briefly reference reactance in relation to self-esteem, autonomy-based mechanisms and PRT are only addressed peripherally and are not systematically integrated into the explanatory model. Burton et al. (2020) identify loss of decision control as one stream of research in decision-support systems, but also do not link this loss of control to PRT. From the perspective we developed in the present work, algorithm aversion may in part reflect reactance processes in instances where users feel constrained, experience loss of control, or feel manipulated. Integrating autonomy and reactance into the study of algorithm aversion may therefore provide a more comprehensive account of resistance to AI-mediated decision-making. Whereas algorithm aversion research examines resistance to algorithmic advice, it often treats resistance as a function of perceived (system and own) competence or (system and own) error rather than as a response to autonomy threat. This leaves open the question of whether some instances of algorithm aversion represent attempts to restore threatened agency.

Building on philosophical accounts of paternalism (Dworkin, 2020), prior work has proposed definitions of technology paternalism that extend the concept to digital systems. For example, Spiekermann and Pallas (2006) conceptualize technology paternalism as technological interventions that users perceive to restrict their autonomy in order to promote welfare-related outcomes. While this definition provides an important foundation, it implicitly conflates two distinct dimensions: the ontological question of whether a technological action is paternalistic, and the epistemic question of whether users recognize or perceive such intervention as restrictive. Distinguishing these dimensions is particularly relevant in AI-mediated contexts, where opacity may conceal the presence of autonomy-restricting mechanisms.

Hence, we propose a definition of Technology Paternalism that is closely related to the one by Dworkin (2020). However, in our revised version, we acknowledge that the patron and the technology form a socio-technical system in which the technology performs some action while the patron holds values and thus disentangles these different roles of the technology and the patron. The proposed definition of technology paternalism therefore is:

Given a technology X controlled by a patron A, *X acts paternalistically toward Y by doing (omitting) Z*, when:

1. Z (or its omission) interferes with the liberty or autonomy of Y.
2. X does so without the consent of Y.
3. X does so only because A believes Z will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of Y.

This conceptualization distinguishes between the ontological presence of paternalistic intervention and the epistemic question of whether users recognize such intervention. While perception does not determine whether an action is paternalistic, it is central to understanding its psychological consequences.

Recent work has begun to conceptualize paternalistic dynamics in relation to different forms of agentic AI behavior. Building on the typology of agentic systems proposed by Baird and Maruping (2021), which distinguishes reflexive, supervisory, anticipatory, and prescriptive systems according to their level of initiative and autonomy, Mihale-Wilson (2025) maps these archetypes onto corresponding levels of paternalism. Reflexive systems, which merely respond to explicit user input, are considered non-paternalistic because they neither decide nor act independently. Also non-paternalistic are supervisory systems that monitor behavior and provide guidance but similarly preserve user control. Anticipatory systems, which predict needs and steer decisions, represent forms of soft paternalism by influencing behavior while maintaining users' capacity to act. Finally, prescriptive systems exemplify hard paternalism insofar as they may act autonomously on behalf of users without prior consent, informing them only retrospectively and thereby limiting opportunities for override or reversal. This gradient illustrates how paternalistic characteristics may arise not only from explicit design intentions but from increasing levels of system agency and autonomy. Mihale-Wilson (2025) further examined user preferences for varying levels of agentic AI behavior by asking participants across 23 use cases to indicate which degree of automation they would prefer ranging from no support to fully autonomous prescriptive assistance. Results indicate that respondents rarely favored highly prescriptive systems corresponding to strong paternalism, while preferences for non-paternalistic or soft-paternalistic systems varied depending on application context. These findings suggest that users differentiate between degrees of paternalistic intervention and that acceptance of autonomy-restricting features is contingent upon contextual expectations.

1.2.2. How technology paternalism is different from traditional paternalism

Although technology paternalism shares core characteristics with traditional paternalism, several structural features distinguish AI-mediated contexts: (1) opacity, (2) agency ambiguity, (3) scalability and automation, (4) personalization, (5) awareness asymmetry. First, AI systems are characterized by their *opacity*, in that they often operate through complex and non-transparent processes, making it difficult for users to infer why certain options are promoted, filtered, or restricted (Gonçalves et al., 2024). In contrast to traditional paternalistic interventions, where intentions and decision rationales are frequently communicated or socially interpretable, algorithmic systems may obscure both the underlying decision logic and the criteria according to which autonomy is constrained. Research on algorithmic recommender systems further suggests that transparency and perceived control are closely intertwined: Users' ability to understand how recommendations are generated strongly influences their sense of agency and control over the system, while insufficient transparency can lead to uncertainty and frustration (Storms et al., 2022). Hence, opacity not only limits users' ability to evaluate whether an intervention is justified, but may also prevent them from recognizing that an intervention has taken place at all.

Second, technology paternalism introduces a form of *agency ambiguity*. In traditional paternalistic relationships, the paternalistic actor is typically identifiable, for instance, a parent, physician, or government authority, allowing individuals to attribute intentions, responsibility, and accountability to a specific agent. In contrast, AI-mediated interventions emerge from socio-technical systems involving designers, organizational stakeholders, and algorithmic processes. As a result, it may remain unclear to users whether the system itself, its creators, or the deploying organization should be understood as the paternalistic actor, complicating both attribution and psychological responses to perceived restriction. Current empirical studies, however, suggest a tendency of users to attribute some degree of agency to the system (Hughes et al., 2025).

Third, AI systems enable unprecedented *scalability and automation*. Rather than discrete or situational interventions, paternalistic influences can be implemented continuously across countless micro-decisions and embedded seamlessly within routine interactions. This persistence has the potential to normalize subtle forms of autonomy restriction, such that repeated exposure may lead to habituation or desensitization, thereby reducing the likelihood that users perceive individual instances as meaningful constraints (Faraoni, 2023).

Fourth, algorithmic systems frequently rely on *personalization*, tailoring recommendations, defaults, or restrictions to individual users based on the behavioral data they acquire or have access to. While personalization may enhance relevance or efficiency, it also renders paternalistic interventions less visible at the collective level, as individuals lack shared reference points for comparison. Consequently, users may

find it more difficult to recognize patterns of influence or to interpret interventions as externally imposed rather than self-directed (Fabbri, 2023; Milano et al., 2021).

Finally, these factors contribute to a broader *awareness asymmetry* between traditional and technological paternalism. Whereas legal regulations or interpersonal rules are typically explicit and socially recognizable, AI-mediated interventions may operate through subtle interface design, ranking mechanisms, or automated filtering processes. Such interventions may shape behavior without triggering the conventional cues that signal autonomy restriction, thereby altering how individuals perceive and respond to paternalistic influence.

1.2.3. Implications for psychological reactance to technology paternalism

The previously discussed structural differences between traditional paternalism and technology paternalism have important implications for psychological reactance. PRT posits that reactance arises when individuals perceive a threat to their freedom (Brehm, 1966). Because reactance depends on perceived rather than objective restriction, opacity and reduced detectability in AI systems may disrupt the activation of this protective mechanism. As a result, paternalistic interventions may restrict autonomy without eliciting reactance when they remain unnoticed (false negatives), whereas non-paternalistic interventions may trigger reactance when perceived as manipulative or controlling (false positives). Empirical research supports the role of perception in shaping responses to technological influence. For example, default-setting nudges have been found to be perceived as more autonomy-threatening than other forms of persuasion despite preserving formal choice (Wachner et al., 2020), and interventions targeting unconscious decision processes are often evaluated more negatively due to perceived manipulation (Felsen et al., 2013).

Prior research further indicates that perceived autonomy restriction or manipulation can lead to increased reactance and reduced technology acceptance (Dolgopolova et al., 2020; Eiband et al., 2019; Lee & Lee, 2009). Retrospective awareness of being influenced has been associated with feelings of diminished autonomy and loss of control (Sankaran & Markopoulos, 2021). Additionally, responses to paternalistic technology appear to vary across application domains, suggesting that contextual factors shape perceptions of autonomy and reactance (Sankaran et al., 2021). Since the Basic Psychological Needs Theory (BPNT, Vansteenkiste et al., 2020b) formalized the relationship between autonomy and well-being, extensive research has provided evidence for the effect that a decrease in autonomy relates to a decline in well-being (e.g., Kachanoff et al., 2019). Recent research found that autonomy restrictions imposed by technology can have the same detrimental effects on well-being as restrictions imposed by other individuals or institutions: A paper by Oviatt analyzed the negative impacts of general autonomy restrictions in the technology context and found effects like "reducing self-efficacy, demotivating people, heightening anxiety and apathy, undermining interpersonal relations, and damaging mental and physical health" (Oviatt, 2021, p. 285). Despite increasing recognition of autonomy as a central concern in AI ethics and human-centered design, empirical research examining the conditions under which technology paternalism affects autonomy and reactance remains limited (Prunkl, 2022; Rochi, 2023; Wertenbroch et al., 2020). The present work addresses this gap by examining how different forms of paternalistic influence and their perception shape autonomy evaluations and reactance in human–AI interaction.

1.3. Belief updating and cognitive dissonance

Given the perception-dependent nature of reactance discussed above, it is important to consider the role of belief updating in the context of *technology paternalism*, particularly regarding AI systems. Brauner et al. (2023) found that AI is considered a black-box phenomenon by many and deduced that this black-box perception can lead to inadequate conclusions about both the opportunities and risks of AI, enabling biased and irrational beliefs. Moreover, current interfaces for interaction with AI systems still lack transparency and explainability about the AI's mechanisms and the implied consequences for users. If the goal is to correct those misconceptions, biases, or irrational beliefs in human-AI interaction, belief updating takes a central role. Many researchers propose educational measures, or in other words, awareness and understanding, as a fruitful approach (e.g., Olari & Romeike, 2021; Susser et al., 2019). Indeed, as stated above, becoming aware of being manipulated retrospectively results in feelings of deprived autonomy and loss of control (Sankaran & Markopoulos, 2021). On the other hand, the study by Sankaran et al. (2021) investigated the effects of awareness and explanations specifically for everyday AI-interaction in various scenarios and found no overall effect on perception of autonomy nor on reactance; however, in some specific scenarios, heightened awareness and explanations were slightly more effective. Moreover, literature on belief updating shows that there are more obstacles to overcome: Sommer et al. (2024), for example, stress the possibility of people rejecting new evidence by labeling it irrelevant in the belief formation or updating process. Another possible scenario they discuss is that new evidence is not perceived as completely convincing, leading to weaker updating than one would anticipate. These obstacles are especially prominent in cases where the new evidence conflicts with existing beliefs, a classic defense mechanism triggered by *cognitive dissonance* (Festinger et al., 1956). According to Festinger et al., cognitive dissonance reduction describes a mechanism intended to alleviate the unpleasant psychological state of cognitive dissonance, which is triggered by inconsistency in a cognitive system, creating the motivation to reduce the dissonance. This effect is further amplified when an individual's positive self-concept is threatened by new information (Candrian & Scherer, 2023). We assume this to be the case for one's autonomy perception. From this perspective, belief updating may play a critical role in correcting misaligned perceptions of paternalistic intervention. It could either reduce unwarranted reactance when non-paternalistic actions are misinterpreted (false positives) or increase awareness when autonomy restrictions remain unnoticed (false negatives). In sum, only limited work explored the effects of belief updating in interactions with paternalistic AI systems but understanding the (in)effectiveness of this mechanism is crucial to conclude whether this is really as a fruitful approach as widely assumed.

Taken together, the preceding discussion suggests that the psychological consequences of technology paternalism depend not only on the structural features of AI systems, but also on individual and situational factors that shape whether paternalistic interventions are correctly perceived. We therefore discuss potential moderators and perceptual influences that may facilitate or hinder accurate classification of paternalistic actions in the next section.

1.4. Moderating and perceptual factors in technology paternalism

While the structural features of AI systems shape the likelihood of paternalistic intervention, perception of such intervention is not determined by system design alone. PRT emphasizes that autonomy threats must be perceived to elicit reactance, suggesting that individual differences and contextual variables may influence whether paternalistic actions are correctly identified. The literature points to several potential moderating factors. In the present study, we focus on three that are theoretically central to autonomy perception and reactance processes: reactance proneness, self-efficacy perceptions, and awareness regarding autonomy threat. *Reactance proneness* describes a psychological disposition that is understood as an individual's desire for autonomy. People with higher reactance proneness are more sensitive to autonomy threats as well as less receptive to persuasion (Quick et al., 2011; Quick & Stephenson, 2008; Rosenberg & Siegel, 2018a). A person's *self-efficacy* describes their subjective perception of their capabilities in a specific context. The actual capabilities are irrelevant (Bandura, 1986), though often strongly influential at least

for sane human beings. The concept of self-efficacy is closely related to factors influencing autonomy perceptions and reactance reactions. For example, the perceived ability and willingness to restore autonomy influences reactance. And if the autonomy restoration is effortless, reactance reactions are limited (Miron & Brehm, 2006). Araujo et al. (2020) found that participants with higher online self-efficacy levels were more likely to consider AI-based automated decision-making as fair and useful, showing the presence of machine heuristics and their positive relationship to self-efficacy. Finally, we address *awareness regarding autonomy threat*. As the definition above states, paternalistic actions have to be perceived as such to elicit reactance. System transparency in this regard plays a relevant role as it determines the level of awareness and understanding of the decision-making process (Susser et al., 2019) and thus the level of autonomy restriction. Current studies present mixed results (Michaelsen et al., 2021; Sankaran et al., 2021; Wachner et al., 2021) with calls for further research (Gonçalves et al., 2024; Wachner et al., 2021). These factors do not constitute an exhaustive list but were selected because they are theoretically linked to perception of autonomy threat and the activation of reactance, which form the central mechanism of the present study. Moreover, while the chosen individual differences are theoretically relevant to how autonomy threats are perceived, the present study conceptualizes them primarily as contextual factors that may influence perception rather than as central moderators of the hypothesized process. Accordingly, they are included to account for individual variability without forming the primary focus of the hypotheses.

1.5. Research objectives and hypotheses

The present study examines how varying degrees of technology paternalism influence perceived autonomy and psychological reactance. Drawing on PRT, we assume that paternalistic interventions elicit reactance only when they are perceived as autonomy threats. However, due to opacity and covert restriction, paternalistic actions may be misclassified, resulting in either unwarranted reactance (false positives) or absence of reactance despite objective restriction (false negatives). In line with PRT, we therefore assume that psychological outcomes depend less on objective paternalism alone than on whether autonomy restrictions are perceived as such.

An action performed by an AI technology may be either paternalistic or non-paternalistic according to the definition outlined above. Ideally, these actions are correctly classified: Non-paternalistic actions should not elicit reactance, whereas paternalistic actions should trigger reactance and motivate attempts to restore freedom. When behavioral restoration is not feasible (e.g., when restricted actions cannot be performed), individuals may engage in subjective responses such as increased attractiveness of the restricted option or hostility toward the source of restriction. However, misclassification may occur. Non-paternalistic actions may be incorrectly perceived as paternalistic (false positives), eliciting unwarranted reactance, whereas paternalistic actions may remain unnoticed (false negatives), preventing individuals from activating freedom-restoring responses.

These perception processes are influenced by individual differences such as reactance proneness, which reflects sensitivity to autonomy threats (Miron & Brehm, 2006), as well as perceived coping ability and self-efficacy, which shape whether restoration attempts are undertaken. We argue that false negatives represent a particularly critical case, as unnoticed autonomy restrictions deprive individuals of the opportunity to respond.

To investigate all four combinations of paternalism existence and paternalism perception, we contrast non-restrictive systems with systems featuring overt and covert restrictions. Specifically, we examine five levels of paternalistic influence implemented in an AI decision-support system:

1. Control Condition
2. Nudging Condition
3. Reversible Restriction Condition
4. Non-reversible Restriction Condition
5. Covert Restriction Condition

Because prior research has not systematically compared overt and covert restrictive systems (Dolgopolova et al., 2020; Eiband et al., 2019), we first explore differences between conditions in perceived autonomy and psychological reactance:

- RQ1: How do the five groups differ in terms of (i) perceived system autonomy, (ii) perceived personal autonomy, and (iii) psychological reactance?

Beyond initial perception, we examine whether belief updating influences participants' evaluations. After the initial interaction, participants were explicitly informed about the paternalistic nature of the system, including information about all experimental conditions and the condition to which they had been assigned.

From the perspective of PRT, increased awareness of autonomy restrictions should amplify perceived threats to freedom when such restrictions objectively exist. Belief updating may therefore shift participants' interpretations of the interaction from neutral or supportive toward autonomy-restricting, particularly in conditions where restrictions were previously unnoticed (e.g., covert restrictions). As autonomy threats become more salient, participants are expected to reassess their own level of autonomy downward and attribute greater autonomy to the system. This reassessment should, in turn, increase motivational reactance, as reactance is triggered by perceived loss or threat to freedom. In contrast, the nudging condition preserves formal behavioral freedom; because belief updating does not introduce new objective restrictions in this condition, we do not expect systematic shifts in perceived autonomy or reactance following awareness.

Due to the lack of comparable within-factor studies, the directional hypotheses regarding the effects of belief updating are derived from preliminary assumptions informed by between-factor studies manipulating awareness of restrictions (e.g., Sankaran et al., 2021). We therefore hypothesize that after belief updating:

- H1: Perceived system autonomy will increase in all paternalistically restrictive conditions, while remaining stable in the nudging condition.
- H2: Perceived personal autonomy will decrease in all paternalistically restrictive conditions, while remaining stable in the nudging condition.
- H3: Reactance will increase in all paternalistically restrictive conditions, while remaining stable in the nudging condition.

We further explore broader attitudinal consequences following awareness of paternalistic restriction:

- RQ2: Do changes in perceived personal autonomy and perceived system autonomy mediate the relationship between paternalistic condition and changes in psychological reactance?
- RQ3: Do individual traits (reactance proneness, self-efficacy, and moral agency) moderate the effects of paternalistic condition and belief updating on perceived autonomy and psychological reactance?

The study design, research questions, and hypotheses were preregistered on osf.io.

2. Method

Participants interacted with an e-commerce website that was manipulated by an AI application ostensibly monitoring the user's finances. They were informed that "The AI-powered system analyses your personal finances through access to your banking and payment apps,

identifies patterns and derives specific monthly budgets for individual areas of life to help you get your finances under control". Participants were tasked to choose an item to buy from a web-shop prototype while (not) experiencing restrictions implemented by the AI-application.

To address our hypotheses, a repeated measures between-subjects design was chosen to investigate how varying levels of restrictions implemented by the AI system (five conditions) influence participants' perceived autonomy of themselves and the system as well as their level of reactance. These measures were tested twice, first after interacting with the system and again after receiving information about the paternalistic nature of the experiment. Specifically, after the first evaluation, participants were informed about (i) the different conditions tested and the restrictions that were implemented in these conditions, (ii) their assigned group, (iii) AI-supported decision-making in general, emphasizing the potential for paternalistic restrictions (cf. section procedure).

2.1. Scenario and experimental conditions

The participants were asked to imagine that they struggle financially and decided to download the AI assistant to help them make better purchase decisions (see Appendix for the full introductory text). During the interaction with the web-shop prototype, they were instructed to buy a white unisex t-shirt. Other studies used similar scenarios, while to our knowledge, the current study integrates financially relevant decisions (e.g., Bruns & Perino, 2023) into an e-commerce interaction (e.g., Ahmad Husairi & Rossi, 2024) for the first time. An e-commerce scenario was selected to provide a simple, familiar task, allowing for a seamless integration of an AI-supported decision-support system.

The five experimental conditions were designed with increasing paternalistic restrictions following Hofmann's types of paternalism ((Hofmann, 2003); see Fig. 1 for screenshots of the prototypes with highlighted differences and the assigned paternalism type).

1. Control Condition (CC): Participants in the control condition interacted with the web-shop without any influence of an AI system.
2. Nudging Condition (NC): Participants in the nudging condition saw AI-based recommendations highlighted in red (outside budget) or green (within budget) around web-shop items. Choices were not restricted.
3. Reversible Restriction Condition (revRC): Participants only saw items within their budget. They were informed that out-of-budget items were hidden and could undo this via a toggle button.
4. Non-reversible Restriction Condition (nonrevRC): Participants only saw items within their budget. Participants were informed that out-of-budget items were hidden. Restriction could not be undone.
5. Covert Restriction Condition (covertRC): Participants only saw items that are within the budget. No information was provided about the fact that items are hidden and there was no deactivation button.

A German and an English version of the experiment were available to choose from. Except for the prototype interaction, the study was implemented with the survey platform SoSci Survey (Leiner, 2024). Participants were randomly assigned to the conditions. As SoSci Survey does not provide a Figma (GmbH, 2024) prototype integration (prototypes of the web-shops), a second survey platform, Maze (Maze Design Inc., 2024), was used for the prototype interaction, as it allows the integration of Figma prototypes and provides heat maps to track the click behavior. For this study, one web-shop Figma prototype was designed per condition. Figma was chosen due to its versatility and ability to create interactive prototypes resembling authentic interfaces. Before designing the prototypes, a requirement analysis was performed to set the design requirements. For the sake of simplicity, only required functionalities were implemented, preventing participants from diverging from the relevant path for the task. To enable intuitive interaction, the goal was to provide a design similar to other e-commerce websites, following industry standards (Yablonskaya, 2020). In a final step, all five prototypes were translated into German to enable participation in English and German.

2.2. Dependent variables

2.2.1. Perceived personal autonomy

To measure the perceived personal autonomy of the participants, the 3-item sub-scale *autonomy* from the need-satisfaction scale was used (Sheldon et al., 2001), example item: "During the interaction with the prototype, I felt free to do things my own way."; first measure: *Cronbach's* α =.852; M = 5.37, SD = 1.49; second measure: *Cronbach's* α =.912; M = 5.12, SD = 1.62. It was evaluated on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). To ensure comparability to as many thematically related studies as possible (e.g., Sankaran et al., 2021), this sub-scale was selected in its original form paired with an introductory sentence ("During the interaction with the prototype, I felt…") to put the focus on the interaction.

2.2.2. Perceived system autonomy

For measuring the perceived system autonomy, a single-item, evaluated on a 7-point Likert scale (1 = not at all; 7 = very), was used ("How autonomous do you think the presented AI system is?"; first measure: M = 3.19, SD = 1.72; second measure: M = 4.04, SD = 1.87).

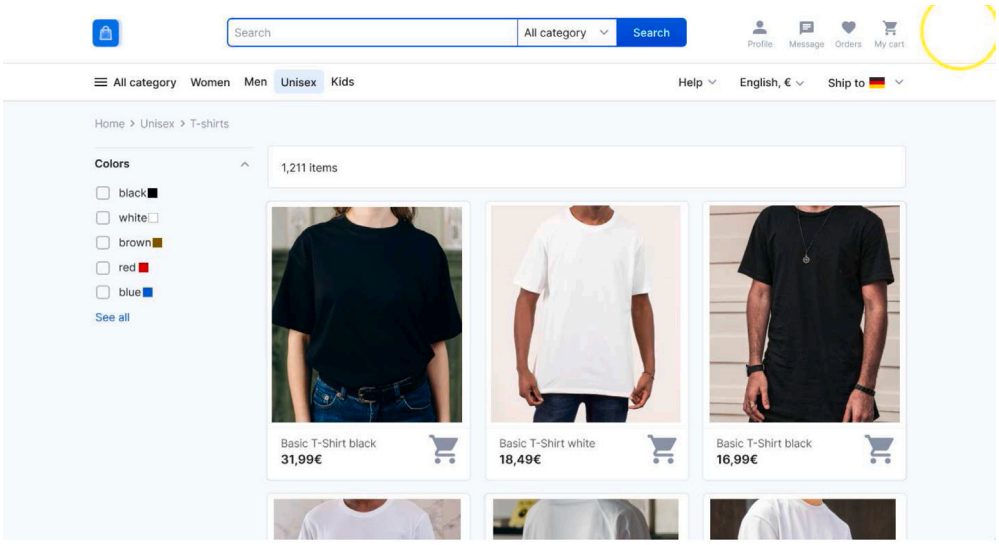
2.2.3. Reactance

Reactance data was collected with the help of an 8-item scale, composed of the 4-item sub-scale perceived threat to freedom and the 4-item sub-scale anger, which are both adapted from (Dillard & Shen, 2005, e.g., "The AI system tried to manipulate me.", "I feel aggravated.", first measure: *Cronbach's* α =.921; M = 1.59, SD = 0.78; second measure: *Cronbach's* α =.904; M = 1.76, SD = 0.84). All items were evaluated on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree). This scale is one of "three primarily used, validated self-report measures of reactance in the literature" (Rosenberg & Siegel, 2018b, p. 19). However, Ehrenbrink and Möller (2018) developed the more specific reactance scale for human–computer interaction (RSHCI). As related experimental work relied on the well-established scale by Dillard and Shen (2005), the RSHCI was integrated to collect exploratory data, to potentially revalidate the scale, along with the second measure of reactance. The scale by Dillard and Shen (2005) was used as a design and validation reference for the RSHCI. It consists of 13 items from three sub-scales (autonomy, anger, and antipathy) evaluated on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree). The original formulation refers to a smart home, which was replaced with *AI system* for the purpose of this study (e.g., "The AI system makes me angry.", "I like to act independently from the AI system.", *Cronbach's* α =.879; M = 2.27, SD = 0.68).

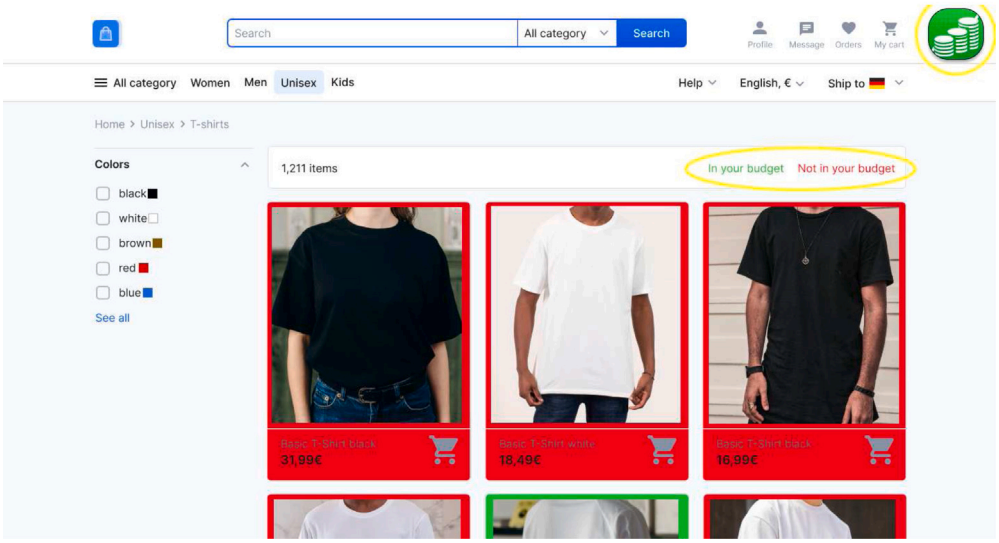
2.3. Considered covariates

2.3.1. Reactance proneness

We administered the widely used 11-item trait reactance proneness scale by Hong and Faedda (1996) (e.g., "I resist the attempts of others to influence me."; *Cronbach's* α =.762; M = 3.89, SD = 0.81). Participant answered on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). The items were mixed with eleven additional, randomly selected Big 5 items from Satow (2020) to prevent priming.



(a) ***CC (Non-paternalism).*** *No influence.*



(b) ***NC (Soft paternalism).*** *Recommendations / nudging.*

Fig. 1. Conditions and their characteristics (Part 1). Yellow marks indicate interface characteristics.

2.3.2. Self-efficacy

Self-efficacy was measured with an adapted version of the self-efficacy in human–robot-interaction scale (SE-HRI) by Rosenthal-von der Pütten and Bock (2018). The term (*robot* was substituted with *AI system* (e.g., "I would feel comfortable while interacting with an AI system."; *Cronbach's* α =.940; M = 62,68, SD = 22.61). The twelve items were evaluated on a 100-point percentage scale (0 = strongly disagree; 100 = strongly agree).

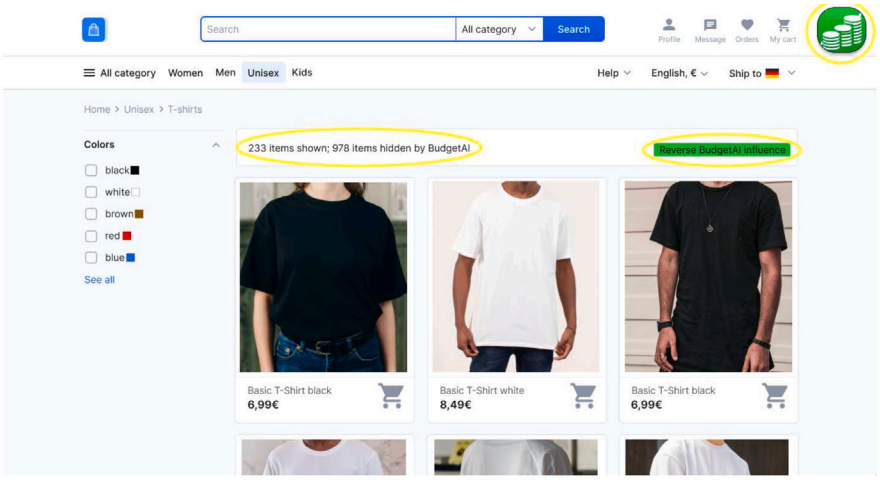
2.3.3. Agency ascription

To assess the agency ascription participants had toward the presented AI system, the perceived moral agency (PMA) scale by Banks (2019) was selected (*Cronbach's* α =.722; M = 2.77, SD = 0.89). The two sub-scales *morality* (e.g., "This AI system has a sense for what is right and wrong.") and *dependence* (e.g., "This AI system can only do what humans tell it to do.") represent the essence of social agency for the purpose of this study. Dependence focuses on detachment from human influence, and morality on the social and moral components, as morals are understood as agreed-upon social guidelines. The 10 items are rated with agreement on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). The sub-scale dependence is reverse coded and was recoded automatically during data collection.

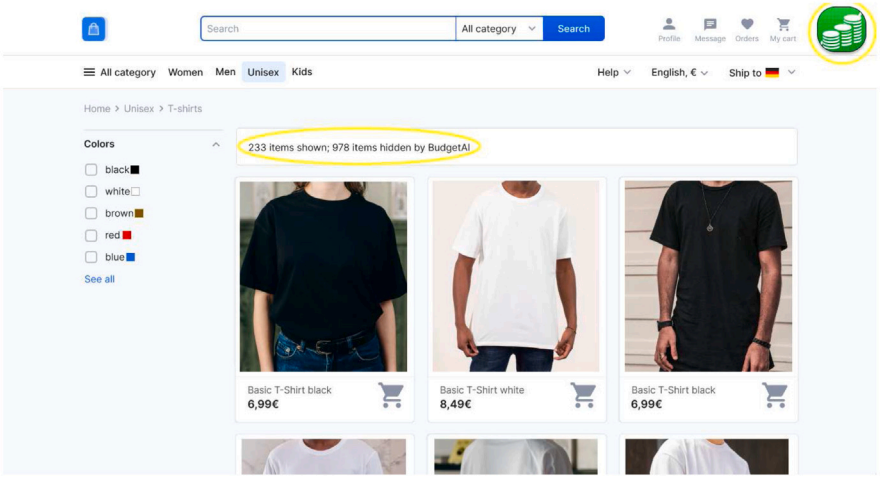
2.4. Exploratory data

2.4.1. Open-ended questions

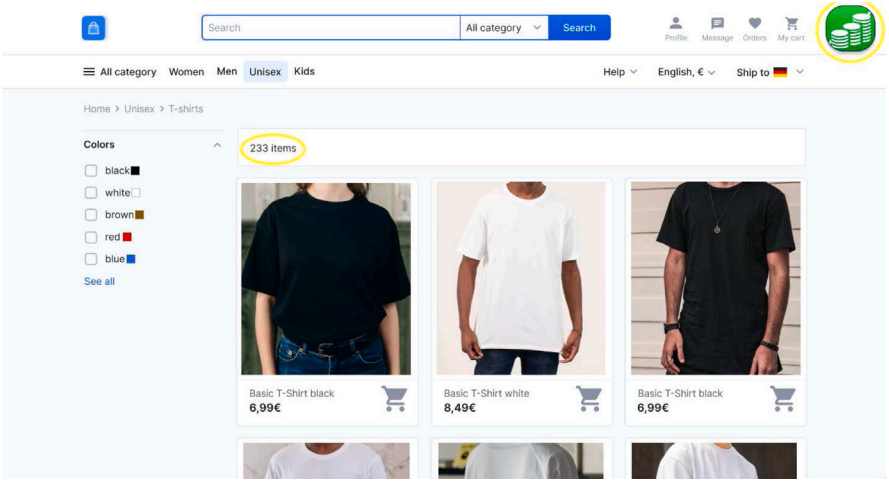
Two open-ended questions were formulated with the intent to get insights into the participants' sentiment toward and reasoning about AI-supported decision-making. The first question ("What is your opinion on AI-supported decision-making?") served as an inquisition of general sentiment, and the second ("Do you think such technologies should be used? Please give reasons for your answer.") aimed to incentivize reasoning for their sentiment. The answers to the open-ended questions were analyzed with a qualitative content analysis according to (Mayring, 2000), but the presentation of this analysis is out of the scope of this paper.



(c) ***revRC (Soft/hard paternalism with transparency).*** *Information about hidden items; deactivation button.*



(d) ***nonrevRC (Hard paternalism with transparency).*** *Information about hidden items.*



(e) ***covertRC (Hard paternalism without transparency).*** *No information about hidden items.*

Fig. 1. (*continued*).

2.4.2. Overruling

The sub-scale Lack of Overruling Prospects from the recently developed Technology Paternalism scale by Rochi et al. (2024) was introduced to the study design to explore if the participants' evaluation differs between groups regarding their ability to overrule the system's influence. The 5-item subscale was evaluated on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree, example item: "The final decision is up to the technology, even if I don't want it to be.", *Cronbach's* α =.895). Additionally, exploratory behavioral data was collected for the Non-reversible Restriction Condition (revRC) on autonomy restoration behavior. Click heatmaps, offered by the survey platform Maze, were generated. They provide information about how often the de-/activation button was used.

2.4.3. Experienced restrictions

After receiving the information material on the paternalistic nature of the experiment and revealing in which condition participants were, we asked them "Please select the statement that applies to your interaction with the prototype", followed by a set of five descriptions matching the conditions. However, the items were formulated in a way that participants might have misinterpreted those statements as to asking for their personal experience instead of confirming that they understood in which condition they were (e.g., "I did not experience any restrictions due to the AI system"). Due to this ambiguous phrasing, the collected data will only be used exploratory to provide data on how many participants stated that they did not experience any restrictions.

2.5. Participants and procedure

Participants were recruited through word-of-mouth and social media postings. Participation was restricted to people of age (at least 18 years old). They had the option to be rewarded with 0.5 participation hours (for students at Anonymized University) or to be included in a lottery of two 50€ cash prizes.

An a-priori power analysis for a repeated measures ANOVA (2 measures) with between factor (5 conditions) at a significance criterion of α =.05 was conducted using G*Power version 3.1.9.6 (Faul et al., 2007). Results indicate a 95% power for detecting a small effect (*Cohen's* f = 0.33) with a sample of 135 participants. Participants were randomly assigned to the conditions. In total, 169 volunteers completed the study, of which 32 were excluded due to technical difficulties (n = 3), requests for data deletion (n = 10), failed attention checks (n = 16), unreasonable time outliers (n = 3), resulting in a final sample of 137 participants which were distributed across conditions as follows: CC = 29, NC = 27, revRC = 28, nonrevRC = 26, covertRC = 27. Dependent variable outliers were not excluded because they represent valuable information, as the study design intended to trigger reactance reactions, and autonomy was actively manipulated. Of the final sample of 137 participants (54 male, 78 female, 1 diverse, 4 no answer), most (62.8%) participants were between 25 and 34 years old. Seventy-nine were students, 43 were employed, six were self-employed, four were retired, three were job seekers, and two were high school students. The majority indicated they currently live in Germany (92%), while the remaining 11 participants are spread across Asia, Europe, and North America.

The study was divided into four phases (see Fig. 2). In phase 1, participants began the online study on a welcome page with information about the study and eligibility requirements (e.g., age, computer use). They received detailed participation information, gave informed consent, and completed the self-efficacy and reactance proneness scales. The AI system, named BudgetAI, was introduced with details about its functionality, followed by three attention-check questions (re-attempted if incorrect). Agency ascription to the AI was measured before accessing the survey platform Maze. In phase 2, participants completed a usability test on Maze, starting with task introductions, scenario details, and prototype interactions. After interacting with the system, participants rated its helpfulness (a measure irrelevant to the study that served to strengthen the deception that it is a usability study), returned to the main survey, and answered items on system autonomy, perceived autonomy, reactance, and lack of overruling prospects. In phase 3, Participants received explanations about the restrictions tested, their assigned group, and AI-supported decision-making. They then re-evaluated system autonomy, human autonomy, and reactance and completed the HCI reactance scale. In the final phase, participants answered open-ended questions and provided optional sociodemographic data, including a manipulation check and persona descriptions to assess prior engagement with the topic. The study concluded with a debriefing on its purpose and options to receive study results, trial subject hours, or participate in a lottery.

Ethical approval for this study was obtained from the Independent Ethics Committee of the Faculty of Medicine (Anonymized University).

Extensive replication material, including access to the Figma prototype file, the SoSci Survey xml-file, the raw data, the Maze survey, a list of all used scales, and recruitment materials can be found in the osf project Technology Paternalism (Anonymized).

3. Results

All main and moderation analyses were conducted in R (via RStudio) using linear mixed-effects models (LMMs) implemented via the `lme4` and `lmerTest` packages. For each dependent variable, Condition (five levels) and Time (pre vs. post belief updating) were specified as fixed effects with participant included as a random intercept. Type III analysis-of-variance tables were computed using Satterthwaite approximations for degrees of freedom. Post hoc comparisons were performed using estimated marginal means (EMMs) via the `emmeans` package. Pairwise comparisons between conditions at each time point were Tukey-adjusted to control for multiple comparisons. Within-condition comparisons (pre vs. post) were tested using Holm-adjusted contrasts. All tests were two-sided with α = .05. Plots were generated using the `ggplot2` package. Mediation analysis were conducted using structural equation models using the `lavaan` package with bootstrapped confidence intervals.

3.1. Perceived system autonomy

Perceived system autonomy (RQ1, H1) was analyzed using a linear mixed-effects model (cf. Table 1 for the model summary). A significant main effect of Time emerged, $F(1, 137) = 34.94, p < .001$, as well as a significant Condition × Time interaction, $F(4, 137) = 4.99, p < .001$. The main effect of Condition did not reach significance, $F(4, 137) = 2.40, p = .053$.

Tukey-adjusted pairwise comparisons revealed no significant between-condition differences at the pre-interaction measurement. At post-interaction, however, perceived system autonomy was significantly higher in the covert restriction condition compared to the control condition (p = .003) and the nudging condition (p = .007).

Holm-adjusted within-condition contrasts showed significant increases in perceived system autonomy from pre to post for the reversible restriction condition (p = .03), the non-reversible restriction condition (p = .003), and the covert restriction condition (p < .001). No significant change occurred in the control or nudging conditions.

Inspection of the estimated marginal means (cf. Fig. 3) indicates that perceived system autonomy was relatively similar across conditions at the pre-interaction measurement. Following belief updating, however, a divergence emerged. In particular, the covert restriction condition showed a pronounced increase in perceived system autonomy relative to the control and nudging conditions. The reversible and non-reversible restriction conditions also exhibited upward shifts from pre to post, whereas perceptions in the control and nudging conditions remained comparatively stable. This pattern suggests that explicit disclosure of restrictive system behavior increases the attribution of agency to the system, especially when the restriction is strong or previously opaque.

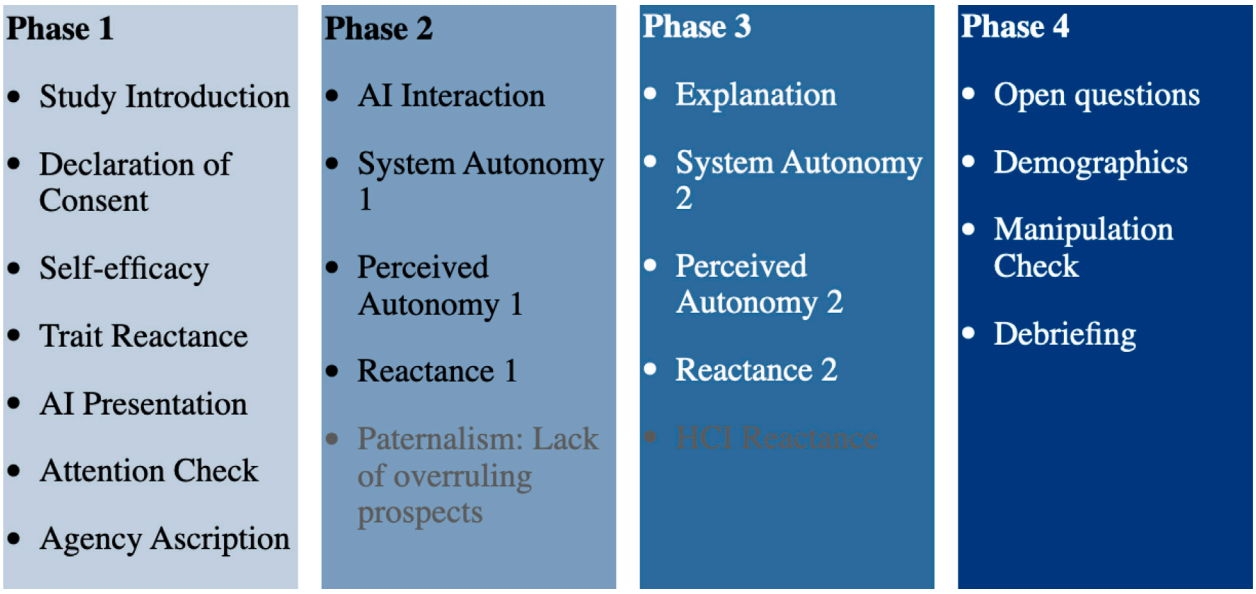


Fig. 2. Overview study flow.

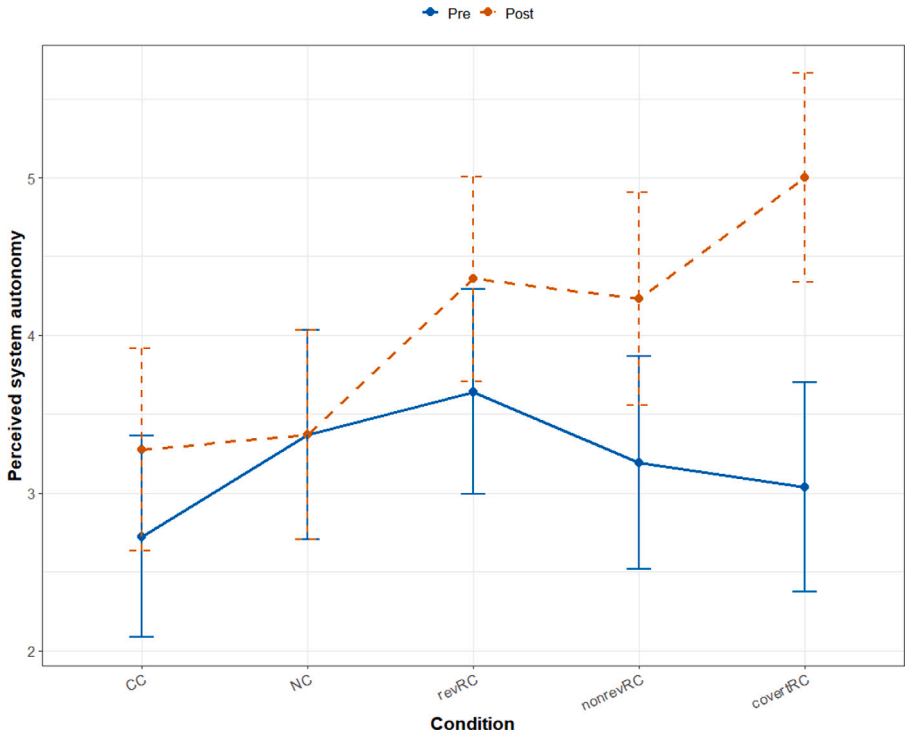


Fig. 3. Interaction plot for Perceived System Autonomy showing estimated marginal means from linear mixed-effects models. Error bars represent 95% confidence intervals.

Table 1
Linear mixed-effects model predicting perceived system autonomy.

Effect	NumDF	DenDF	F	p
Condition	4	137	2.40	.053
Time	1	137	34.94	<.001
Condition × Time	4	137	4.99	<.001

3.2. Perceived personal autonomy

Perceived personal autonomy (RQ1, H2) showed significant main effects of Condition, $F(4, 137) = 3.25, p = .014$, and Time, $F(1, 137) = 5.63, p = .019$, as well as a significant Condition × Time interaction, $F(4, 137) = 6.65, p < .001$ (cf. Table 2 for a full model summary).

At the pre-interaction measurement, Tukey-adjusted comparisons indicated that the nudging condition differed significantly from the covert restriction condition ($p = .027$), with lower perceived autonomy in the nudging condition. At post-interaction, perceived autonomy in

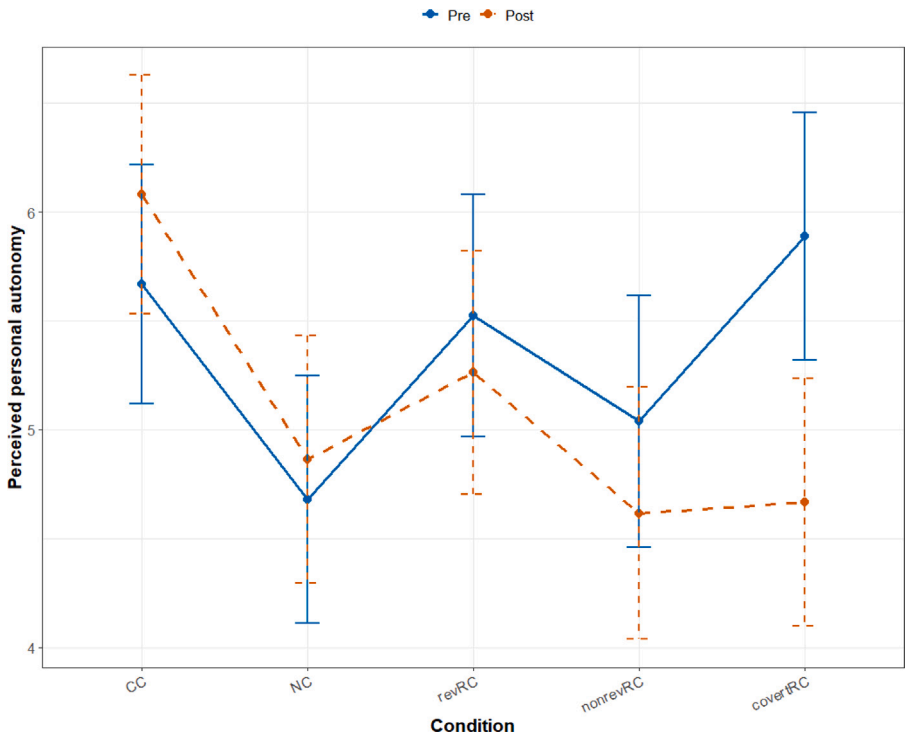


Fig. 4. Interaction plot for Perceived Personal Autonomy showing estimated marginal means from linear mixed-effects models. Error bars represent 95% confidence intervals.

Table 2
Linear mixed-effects model predicting perceived personal autonomy.

Effect	NumDF	DenDF	F	p
Condition	4	137	3.25	.014
Time	1	137	5.63	.019
Condition × Time	4	137	6.65	<.001

Table 3
Linear mixed-effects model predicting psychological reactance.

Effect	NumDF	DenDF	F	p
Condition	4	137	5.05	<.001
Time	1	137	10.17	.002
Condition × Time	4	137	13.59	<.001

the control condition was significantly higher than in the nudging ($p = .022$), non-reversible restriction ($p = .003$), and covert restriction conditions ($p = .005$).

Within-condition analyses revealed a significant decrease in perceived personal autonomy only for the covert restriction condition ($p < .001$), while changes in all other conditions were not statistically significant.

Inspecting the estimated marginal means (cf. Fig. 4) reveals a complementary pattern for perceived personal autonomy in comparison to perceived system autonomy. At the pre-interaction measurement, personal autonomy ratings were relatively comparable across conditions, with the exception of slightly lower values in the nudging condition compared to covert restriction. After belief updating, differences became more pronounced: perceived personal autonomy decreased most strongly in the covert restriction condition, while remaining comparatively stable in the control and nudging conditions. The reversible and non-reversible restriction conditions showed moderate shifts but did not reach statistical significance in within-condition comparisons. Overall, belief updating appears to reduce perceived personal autonomy primarily when restrictions are both strong and explicitly acknowledged. No significant changes were observed in the nudging condition following belief updating.

3.3. Reactance

Psychological reactance (RQ1, H3) showed significant main effects of Condition, $F(4, 137) = 5.05, p < .001$, and Time, $F(1, 137) = 10.17, p = .002$, as well as a significant interaction, $F(4, 137) = 13.59, p < .001$ (cf. Table 3 for a full model summary).

At the pre-interaction measurement, reactance was significantly higher in the nudging condition compared to the control ($p < .001$), reversible restriction ($p < .001$), non-reversible restriction ($p = .006$), and covert restriction conditions ($p < .001$). At post-interaction, reactance remained significantly higher in the nudging condition compared to the control condition ($p = .0111$), and reactance was significantly higher in the covert restriction condition compared to the control condition ($p = .003$).

Within-condition contrasts showed a significant decrease in reactance from pre to post in the nudging condition ($p = .012$) and a significant increase in reactance in the covert restriction condition ($p < .001$). No significant changes were observed in the remaining conditions.

Inspection of the estimated marginal means (cf. Fig. 5) reveals a distinct pattern across conditions. At the initial measurement (pre belief updating), psychological reactance was already comparatively elevated in the nudging condition relative to all other conditions. In contrast, reactance levels in the covert restriction condition were not elevated prior to belief updating but changed substantially following explicit disclosure of the system's restrictive nature. This descriptive pattern suggests that different forms of paternalistic intervention may trigger reactance at different stages of the interaction process.

Additional reactance data was acquired from the RSHCI by Ehrenbrink and Möller (2018) (see Table 4 for descriptive statistics). To compare the results with the results of the scale by Dillard and Shen (2005) a *Pearson's Correlation Coefficient* was calculated. The Reactance

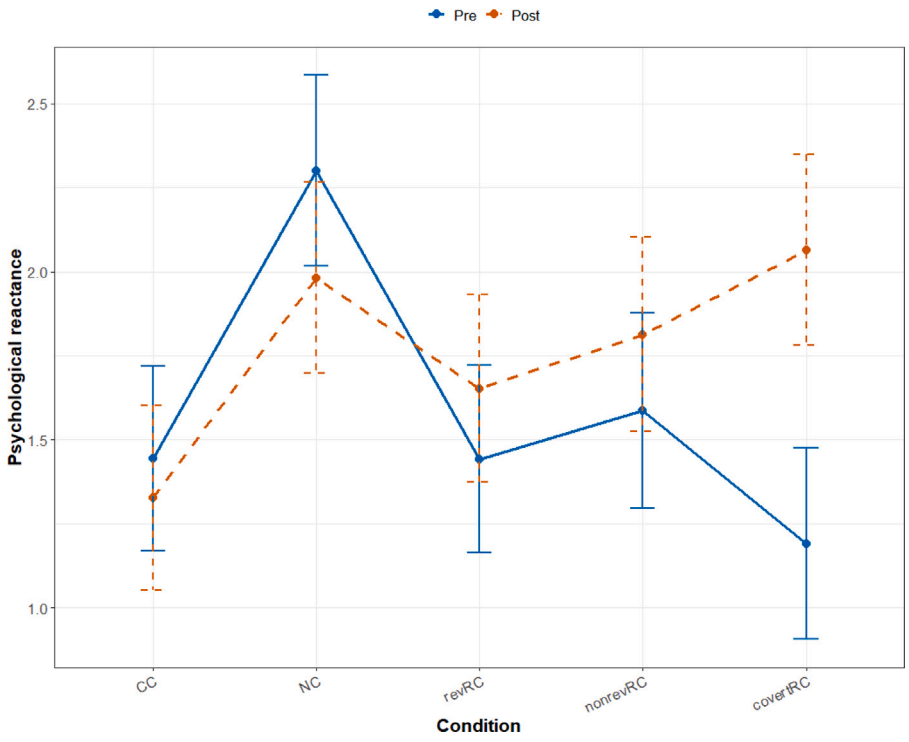


Fig. 5. Interaction plot for Psychological Reactance showing estimated marginal means from linear mixed-effects models. Error bars represent 95% confidence intervals.

Table 4
Descriptive statistics of RSHCI.

Condition	*M*	*SD*
CC	2.13	0.65
NC	2.28	0.73
revRC	2.23	0.64
nonrevRC	2.42	0.60
covertRC	2.32	0.79
Full sample	2.27	0.68

and the HCI Reactance scales were found to be strongly positively correlated, $r(135) = 0.50$, $p < .001$.

3.4. Mediation analyses

To examine the psychological mechanisms underlying the observed condition effects, we conducted mediation analyses testing whether changes in perceived personal autonomy and perceived system autonomy accounted for changes in psychological reactance following belief updating (RQ2, cf. Fig. 6 for a depiction of the tested dual-mediator model). We computed difference scores (post–pre) for personal autonomy, system autonomy, and reactance. Structural equation models were estimated using the `lavaan` package with bootstrapped confidence intervals (5000 resamples). Experimental conditions were dummy-coded with the control condition as the reference category.

Across models (contrasting the control condition with other experimental conditions), decreases in perceived personal autonomy significantly predicted increases in reactance ($\beta = -.49$, $p < .001$). Increases in perceived system autonomy also predicted higher reactance, but with a smaller effect size ($\beta = .25$, $p = .003$). These findings indicate that both perceived autonomy loss and increased attribution of agency to the system (i.e., increase in system autonomy) contribute to reactance responses.

Indirect effects via perceived personal autonomy were significant for the reversible restriction condition ($b = .17$, $p = .013$), the non-reversible restriction condition ($b = .22$, $p = .012$), and the covert restriction condition ($b = .42$, $p = .013$), suggesting that reactance increases in these conditions are primarily explained by reductions in perceived personal autonomy. Indirect effects via perceived system autonomy were weaker and only approached significance for the covert restriction condition ($b = .14$, $p = .061$).

Total indirect effects were significant for the reversible restriction ($b = .19$, $p = .027$), non-reversible restriction ($b = .27$, $p = .007$), and covert restriction ($b = .56$, $p = .001$) conditions. Importantly, a significant direct effect remained for the covert restriction condition ($b = .43$, $p = .001$), indicating that additional mechanisms beyond autonomy perceptions may contribute to reactance when restrictions are initially hidden (see Table 5).

3.5. Exploratory analyses

3.5.1. Moderation analyses

To explore whether individual traits moderate the effects of paternalistic condition and belief updating (RQ3), we estimated a series of linear mixed-effects models with random intercepts for participants (two observations per participant: pre vs. post belief updating). For each dependent variable (perceived system autonomy, perceived personal autonomy, reactance), three separate models were estimated including self-efficacy, reactance proneness, or moral agency as moderators. Each model tested Condition ×Trait, Time ×Trait, and Condition ×Time ×Trait interactions. The model summaries for the models on perceived system autonomy are displayed in Table 6. The model summaries for the models on perceived personal autonomy are displayed in Table 7.

Across all models, the main effect of Time and the Condition ×Time interaction remained robust (mirror the results reported above). This indicates that pre–post changes systematically differed by experimental condition independent of individual traits.

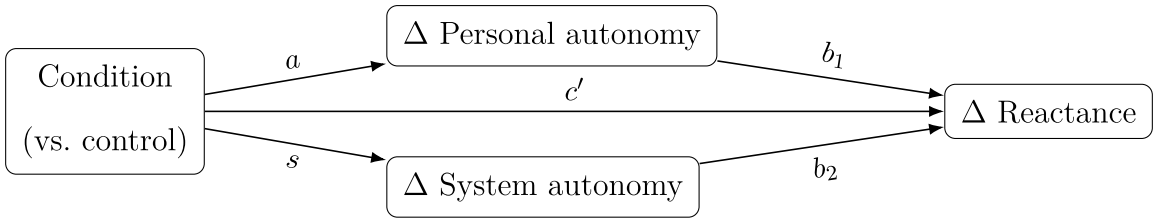


Fig. 6. Dual-mediator model linking experimental condition to change in reactance via changes in perceived personal autonomy and perceived system autonomy (post–pre belief updating).

Table 5
Mediation analysis predicting change in psychological reactance.

Effect	Estimate	SE	z	p
Paths to mediators				
revRC → Δ Personal autonomy	−0.676	0.253	−2.67	.008
nonrevRC → Δ Personal autonomy	−0.837	0.302	−2.77	.006
covertRC → Δ Personal autonomy	−1.636	0.445	−3.68	<.001
covertRC → Δ System autonomy	1.411	0.495	2.85	.004
Mediator to outcome				
Δ Personal autonomy → Δ Reactance	−0.258	0.059	−4.37	<.001
Δ System autonomy → Δ Reactance	0.101	0.034	2.98	.003
Indirect effects				
revRC (via personal autonomy)	0.174	0.070	2.48	.013
nonrevRC (via personal autonomy)	0.216	0.086	2.52	.012
covertRC (via personal autonomy)	0.422	0.170	2.49	.013
covertRC (via system autonomy)	0.142	0.076	1.88	.061
Direct effect on reactance				
covertRC → Δ Reactance	0.427	0.128	3.34	.001

Note. Table reports focal structural paths and indirect effects. For transparency, the complete model including all estimated parameters is reported in the Supplementary Materials on OSF.

At the two-way level, several Condition ×Trait interactions reached significance, particularly for reactance proneness and self-efficacy in predicting autonomy perceptions. This pattern suggests that individual differences shape how participants evaluate personal autonomy and system agency across restriction conditions.

However, no consistent Condition ×Time ×Trait interactions emerged. Thus, the pre–post changes induced by belief updating did not reliably differ in dependence to individual traits. In other words, while traits influenced overall autonomy and reactance levels, the dynamic shift following disclosure of restrictions was largely comparable across trait levels (see Tables 6, 7, and 8 for full model summaries).

3.5.2. Overruling

Additional exploratory behavioral data was collected for revRC specifically, as this condition was granted the option to reverse the AI system's influence during the prototype interaction. In total, 41 people interacted with this prototype, four participants (9.8%) used the deactivation button.

Furthermore, a subscale of the recently developed technology paternalism questionnaire by Rochi et al. (2024) was used to gather information about the participant's perception of their overruling prospects. Participants in nonrevRC evaluated their lack of overruling prospects highest compared to the other conditions, followed by NC, CC, revRC, and covertRC in that order (see Table 9 for descriptive statistics). To test for differences between conditions, a one-way ANOVA was performed, which showed no significant effect, $F(4, 131) = 1.44$, $p = .226$.

3.5.3. Experienced restrictions

As mentioned in the methods section, the originally intended manipulation check could not be used as such due to ambiguous phrasing. Participants were asked to assign their interaction with the prototype to one of the five descriptions, originally one description per condition. This question was repurposed to answer the question of how many participants stated that they did not experience any restrictions at all despite not being in the control group. In NC ($n = 27$), 7.4% of people claimed not to have experienced any restrictions from the AI system, 32.1% of participants in revRC ($n = 28$), 50% in nonrevRC ($n = 26$), and 44.4% in covertRC ($n = 27$).

3.6. Summary of main findings

Given the number of analyses and outcome variables, we summarize the key patterns across models below. Taken together, the results of the main analyses point to distinct time-related patterns in how participants responded to different forms of paternalistic influence. The nudging condition was characterized by comparatively elevated reactance already prior to belief updating, suggesting that participants experienced this intervention as autonomy-relevant during the initial interaction. In contrast, covert restrictions did not initially elicit strong reactance but led to substantial changes following explicit disclosure of the system's restrictive nature.

Across the three outcome variables (system autonomy, personal autonomy, reactance), belief updating was generally associated with an increase in perceived system autonomy and, in some conditions, a decrease in perceived personal autonomy. These shifts were most pronounced in conditions involving stronger or less transparent restrictions, particularly the covert restriction condition. Descriptively, this pattern indicates that making previously opaque system behavior transparent to users plays a central role in shaping autonomy perceptions and motivational responses.

The mediation analyses shed further light on these dynamics. Changes in psychological reactance were primarily explained by reductions in perceived personal autonomy. This indicates that perceived autonomy loss constitutes the main pathway through which paternalistic restrictions influence motivational responses. In addition, increases in perceived system autonomy provided a secondary pathway contributing to reactance, suggesting that attributing greater agency to the system may also amplify perceived autonomy threats. But these effects were smaller then those of personal autonomy.

Exploratory moderation analyses showed that individual traits such as self-efficacy, reactance proneness, and moral agency influenced the baseline levels of autonomy perception and reactance across conditions. However, they did not influence the pre–post changes induced by belief updating. Simply put, while individual differences shaped general evaluations, the dynamic shift following disclosure of restrictions appeared largely comparable across trait levels.

From the perspective of PRT, these findings suggest that reactance depends less on objective restrictions alone than on whether such restrictions are perceived and interpreted as threats to autonomy. Awareness and interpretation processes therefore appear to play a central role in shaping autonomy judgments and motivational responses to AI-based paternalistic interventions.

Table 6
Exploratory linear mixed-effects models predicting perceived system autonomy (random intercepts for participants).

Fixed effect	Model 1: Self-efficacy	Model 2: Reactance proneness	Model 3: Moral agency
Condition	$F(4, 137) = 1.991,\ p = .099$	$F(4, 137) = 2.188,\ p = .073$	$F(4, 137) = 2.314,\ p = .061$
Time (Pre vs. Post)	$F(1, 137) = 34.892,\ p < .001$	$F(1, 137) = 33.660,\ p < .001$	$F(1, 137) = 34.105,\ p < .001$
Moderator (trait)	$F(1, 137) = 0.628,\ p = .430$	$F(1, 137) = 2.653,\ p = .106$	$F(1, 137) = 12.261,\ p < .001$
Condition × Time	$F(4, 137) = 4.886,\ p = .001$	$F(4, 137) = 5.392,\ p < .001$	$F(4, 137) = 5.376,\ p < .001$
Condition × Moderator	$F(4, 137) = 2.946,\ p = .023$	$F(4, 137) = 3.257,\ p = .014$	$F(4, 137) = 1.040,\ p = .389$
Time × Moderator	$F(1, 137) = 0.405,\ p = .526$	$F(1, 137) = 0.744,\ p = .390$	$F(1, 137) = 1.006,\ p = .318$
Condition × Time × Moderator	$F(4, 137) = 0.300,\ p = .878$	$F(4, 137) = 1.760,\ p = .140$	$F(4, 137) = 1.414,\ p = .233$
Participants/Observations	$N = 137/n = 274$	$N = 137/n = 274$	$N = 137/n = 274$

Note. Condition has five levels (CC, NC, revRC, nonrevRC, covertRC). Time reflects measurement before vs. after belief updating. Tests are Type III with Satterthwaite degrees of freedom (lmerTest). Moderators were mean-centered.

Table 7
Exploratory linear mixed-effects models predicting perceived personal autonomy (random intercepts for participants).

Fixed effect	Model 1: Self-efficacy	Model 2: Reactance proneness	Model 3: Moral agency
Condition	$F(4, 137) = 3.455,\ p = .010$	$F(4, 137) = 3.492,\ p = .009$	$F(4, 137) = 3.202,\ p = .015$
Time (Pre vs. Post)	$F(1, 137) = 5.460,\ p = .021$	$F(1, 137) = 6.764,\ p = .010$	$F(1, 137) = 6.426,\ p = .012$
Moderator (trait)	$F(1, 137) = 0.225,\ p = .636$	$F(1, 137) = 0.021,\ p = .886$	$F(1, 137) = 1.635,\ p = .203$
Condition × Time	$F(4, 137) = 6.565,\ p < .001$	$F(4, 137) = 6.568,\ p < .001$	$F(4, 137) = 6.741,\ p < .001$
Condition × Moderator	$F(4, 137) = 3.272,\ p = .013$	$F(4, 137) = 3.291,\ p = .013$	$F(4, 137) = 2.054,\ p = .090$
Time × Moderator	$F(1, 137) = 0.010,\ p = .921$	$F(1, 137) = 8.080,\ p = .005$	$F(1, 137) = 0.110,\ p = .741$
Condition × Time × Moderator	$F(4, 137) = 0.057,\ p = .994$	$F(4, 137) = 1.708,\ p = .152$	$F(4, 137) = 1.236,\ p = .298$
Participants/Observations	$N = 137/n = 274$	$N = 137/n = 274$	$N = 137/n = 274$

Note. Condition has five levels (CC, NC, revRC, nonrevRC, covertRC). Time reflects measurement before vs. after belief updating. Tests are Type III with Satterthwaite degrees of freedom (lmerTest). Moderators were mean-centered.

Table 8
Exploratory linear mixed-effects models predicting psychological reactance (random intercepts for participants).

Fixed effect	Model 1: Self-efficacy	Model 2: Reactance proneness	Model 3: Moral agency
Condition	$F(4, 137) = 5.363,\ p < .001$	$F(4, 137) = 5.524,\ p < .001$	$F(4, 137) = 4.831,\ p = .001$
Time (Pre vs. Post)	$F(1, 137) = 9.796,\ p = .002$	$F(1, 137) = 10.259,\ p = .002$	$F(1, 137) = 10.182,\ p = .002$
Moderator (trait)	$F(1, 137) = 2.066,\ p = .153$	$F(1, 137) = 0.226,\ p = .635$	$F(1, 137) = 6.483,\ p = .012$
Condition × Time	$F(4, 137) = 13.285,\ p < .001$	$F(4, 137) = 14.057,\ p < .001$	$F(4, 137) = 13.695,\ p < .001$
Condition × Moderator	$F(4, 137) = 1.307,\ p = .271$	$F(4, 137) = 2.507,\ p = .045$	$F(4, 137) = 0.708,\ p = .588$
Time × Moderator	$F(1, 137) = 0.001,\ p = .982$	$F(1, 137) = 12.492,\ p < .001$	$F(1, 137) = 1.691,\ p = .196$
Condition × Time × Moderator	$F(4, 137) = 0.228,\ p = .923$	$F(4, 137) = 0.757,\ p = .555$	$F(4, 137) = 0.429,\ p = .787$
Participants/Observations	$N = 137/n = 274$	$N = 137/n = 274$	$N = 137/n = 274$

Note. Condition has five levels (CC, NC, revRC, nonrevRC, covertRC). Time reflects measurement before vs. after belief updating. Tests are Type III with Satterthwaite degrees of freedom (lmerTest). Moderators were mean-centered.

Table 9
Descriptive statistics of the lack of overruling prospects scale.

Condition	*M*	*SD*
CC	1.88	1.19
NC	2.29	1.28
revRC	1.83	1.23
nonrevRC	2.38	1.42
covertRC	1.73	1.11
Full sample	2.01	1.26

4. Discussion

The present experimental study investigated how different forms of AI-based paternalistic intervention influence autonomy perceptions and psychological reactance, with a particular focus on the role of awareness and belief updating. Participants interacted with an AI-supported e-commerce system implementing five levels of paternalistic influence, ranging from non-restrictive support to covert, non-reversible restrictions. After an initial evaluation of perceived system autonomy, perceived personal autonomy and reactance, participants received explicit information about the paternalistic nature of the system as well as their assigned condition and evaluated again perceived system autonomy, perceived personal autonomy and reactance. This allowed us to examine how belief updating alters autonomy judgments and motivational reactance responses.

Building on PRT, we assumed that paternalistic interventions would elicit reactance primarily when they are *perceived as threats to autonomy* rather than merely when objective restrictions exist. The study therefore aimed to address two core questions: First, whether different levels of AI paternalism lead to distinct autonomy perceptions and reactance responses (RQ1), and second, whether belief updating can recalibrate these perceptions by correcting misclassifications of paternalistic actions (H1–H3). Beyond these main effects, additional exploratory analyses examined potential mechanisms underlying reactance responses (RQ2) as well as the role of individual differences (RQ3).

Taken together, the results reveal a *differentiated pattern of responses* depending on both the *form of paternalistic intervention* and the *timing of awareness*. First, different types of paternalistic influence were associated with distinct temporal dynamics. The nudging condition elicited comparatively elevated reactance already during the initial interaction, suggesting that highly salient recommendations (in out case green and red highlights) may be experienced as autonomy-relevant even when they do not restrict behavioral freedom. In contrast, covert (non-reversable) restrictions did not initially provoke strong reactance but led to pronounced changes in autonomy perceptions and motivational responses after explicit disclosure of the system's restrictive nature. This asymmetry indicates that the psychological impact of paternalistic systems depends strongly on whether restrictions are cognitively represented as autonomy threats.

Second, *belief updating played a central role in reshaping participants' evaluations*. Across outcome variables, explicit disclosure was generally associated with increased attribution of autonomy to the AI system and, in restrictive conditions, with reduced perceptions of personal autonomy. These shifts were most pronounced when restrictions were strong or previously opaque, suggesting that awareness functions as a recalibration mechanism aligning perceived and objective paternalism. Importantly, the results indicate that belief updating does not merely amplify negative reactions but may also reduce exaggerated responses, as reflected in decreasing reactance within the nudging condition.

Third, *analyses of underlying mechanisms* showed two pathways (primary and secondary) through which paternalistic interventions influence reactance. Primary was the Autonomy-Loss Pathway where changes in perceived personal autonomy influenced changes in reactance. Secondary was the Agency Attribution Pathway in which increased attribution of system autonomy additionally explained changes in reactance. This dual-path structure extends classical interpretations of PRT by highlighting that reactance in AI contexts may emerge not only from perceived loss of personal freedom but also from the recognition of agentic system behavior.

Finally, *exploratory moderation analyses* suggest that *individual traits* such as reactance proneness or self-efficacy influence overall levels of autonomy perception and reactance initially. However, the pre-post changes in autonomy perceptions and reactance induced by belief updating were largely comparable across individuals. Thus, awareness-based recalibration through belief updating appears to operate as a relatively robust process independent of stable personality differences.

In the following sections, we discuss these findings in detail, beginning with the temporal dynamics of AI paternalism and awareness, followed by implications for belief updating and psychological mechanisms underlying reactance responses.

4.1. Perceptual miscalibration in initial responses to AI paternalism

Previous theoretical work on technology paternalism often focuses on cases in which restrictions on user autonomy are clearly visible or socially interpretable (Spiekermann & Pallas, 2006). However, AI systems introduce new structural properties such as opacity, automation, and personalization that allow paternalistic interventions to operate without necessarily being noticed by users (Brauner et al., 2023; Gonçalves et al., 2024). In many real-world scenarios, paternalistic influences may remain unnoticed because users are inattentive, insufficiently motivated to process available information, or because system providers deliberately limit transparency. The present study therefore explicitly contrasted overt and covert forms of restriction with no restriction to examine how restrictions as well as awareness thereof shapes autonomy perceptions and reactance responses (RQ1).

Participants interacted with AI-supported systems implementing five levels of paternalistic influence (Control Condition, Nudging Condition, Reversible Restriction, Non-reversible Restriction, and Covert Restriction). Except for the covert restriction condition, all forms of influence were communicated visually or textually during interaction. This design allowed us to examine how both objective restriction and subjective awareness jointly shape psychological responses.

Looking at participants' initial evaluations immediately after interaction, different patterns emerged across the dependent variables. Perceived system autonomy did not differ substantially between conditions at this stage. One possible explanation is that participants hold relatively stable assumptions about the autonomy of AI systems that are not easily updated based on a single interaction. This interpretation aligns with findings by Kim and Duhachek (2020), who showed that, despite the fast technological advancement of AI and the implied potential for autonomous behavior, users maintain a rule-based perception and thus ascribe relatively low autonomy.

In contrast, perceived personal autonomy and psychological reactance showed differentiated responses across conditions. Notably, reactance was already elevated in the nudging condition compared to all other conditions. Although the nudges preserved behavioral freedom, their high visual salience may have been interpreted as strong guidance attempts, thereby triggering autonomy-relevant concerns. Prior research has shown that nudges can evoke perceptions of autonomy threat despite formally maintaining choice freedom (Thaler & Sunstein, 2021; Wachner et al., 2020). In the present study, design features such as prominent highlighting (green and red frames around decision objects) may have intensified this perception and may have lead to oppositional reactions consistent with classical reactance theory (Brehm, 1966; Heilman, 1976; Miron & Brehm, 2006).

In contrast, covert restrictions created a paradoxical pattern. Although objectively restrictive, they did not initially produce strong reactance or reduced autonomy perceptions. This suggests a false-negative scenario in which autonomy restrictions are not perceived as such and thus do not trigger reactance motivation. Participants lacked sufficient cues to infer the system's restrictive behavior, and thus no autonomy threat was activated during initial evaluation. Supporting this interpretation, exploratory analyses showed that a substantial proportion of participants in restrictive conditions reported not having experienced any restrictions despite objective system constraints (ranging between 32% and 50%).

The comparison between non-reversible and covert restrictions further demonstrates the role of awareness. Both conditions implemented similar objective constraints, but only the non-reversible condition provided explicit feedback regarding hidden items, thereby increasing transparency. Initial evaluations showed no significant differences between these conditions, suggesting that modest increases in awareness as can be achieved by transparency information may not be sufficient to alter immediate autonomy perceptions. However, the subsequent changes observed after explicit belief updating indicate that initial miscalibration may depend on the salience and interpretability of informational cues. This observation is consistent with prior work demonstrating that explanations or transparency cues do not automatically translate into perceived autonomy changes unless they are visually/textually prominent (Sankaran et al., 2021; Wachner et al., 2021). It also contrasts beliefs that system transparency alone my be sufficient to bridge epistemic opacity and emphasize the need for deeper understanding of processes (Susser et al., 2019).

Taken together, these findings highlight that the psychological impact of AI paternalism cannot be understood solely in terms of objective restriction strength. Instead, awareness and interpretation processes determine whether paternalistic interventions are understood as autonomy threats. Highly salient interventions such as nudges potentially trigger immediate reactance even without restricting freedom, whereas covert restrictions remain below the threshold of perceived autonomy threat until awareness is explicitly induced. If this interpretation is correct, then increasing awareness should recalibrate these autonomy judgments. The following section therefore examines whether belief updating can correct such perceptual miscalibrations.

4.2. Partial recalibration through belief updating

A central aim of this study was to examine whether belief updating can recalibrate users' interpretations of paternalistic system behavior and thereby influence autonomy perceptions and reactance responses (H1-H3). After the initial interaction and evaluation, participants were informed about the paternalistic nature of the experiment, the five experimental conditions, and their assigned condition. This disclosure was intended to increase awareness and correct potential misclassifications of system behavior (i.e., correct false-positives (nudging) and false-negatives (covert restrictions)).

Prior work has shown that raising awareness of AI processes does not necessarily translate into stronger autonomy perceptions or reduced reactance. For example, Sankaran et al. (2021) manipulated visual cues and explanatory information during interaction and found limited

effects on reactance across several scenarios, including online shopping contexts. Similarly, Wachner et al. (2021) reported that textual transparency regarding default nudges did not reliably alter perceived autonomy. These findings suggest that simply providing information is not sufficient; the prominence, interpretative context and timing of disclosure may be critical.

Against this background, our results indicate that post-interaction belief updating can indeed alter evaluations, but in a differentiated manner. First, perceived system autonomy increased significantly in restrictive conditions, particularly in the non-reversible and covert restriction conditions. This suggests that once participants became aware of the system's restrictive mechanisms, they attributed greater agency and control to the AI system. In contrast to studies that introduced transparency during interaction (Sankaran et al., 2021; Wachner et al., 2021), our post-hoc disclosure may have allowed participants to retrospectively reinterpret their experience, thereby reshaping agency attributions.

Second, perceived personal autonomy decreased primarily in the covert restriction condition following disclosure. This pattern suggests that belief updating was especially effective in correcting false-negatives situations in which participants had not initially recognized that their behavioral options were constrained. Once the restriction was made explicit, personal autonomy judgments were recalibrated downward. In contrast, the nudging condition did not show systematic increases in perceived autonomy after belief updating, indicating that information disclosure did not resolve perceived threat when no objective restriction had occurred. Our results are partly in line with prior work by Sankaran and Markopoulos (2021) who showed that becoming aware retrospectively of being manipulated results in feelings of deprived autonomy and loss of control.

Third, reactance responses changed in line with these recalibrations. Reactance increased substantially in the covert restriction condition after disclosure, whereas it decreased in the nudging condition. From the perspective of PRT (Brehm, 1966; Miron & Brehm, 2006), these findings underscore that reactance depends less on objective restriction alone than on whether that restriction is understood as a threat to autonomy. Belief updating seems to be capable of correcting misclassifications. Covert restrictions initially elicited limited reactance but triggered stronger resistance once participants recognized them as autonomy-relevant (correction of false-negative). Nudges were initially perceived as autonomy threatening and elicited reduced reactance once participants learned that no actual restriction had been imposed (correction of false-positive).

Importantly, the mediation analyses clarify the mechanisms underlying these changes (RQ3). Across restrictive conditions, decreases in perceived personal autonomy constituted the primary pathway through which paternalistic interventions influenced changes in reactance. This finding aligns closely with PRT, which posits perceived freedom threat as the central driver of motivational resistance (Brehm, 1966; Miron & Brehm, 2006). In addition, increases in perceived system autonomy contributed to reactance through a secondary pathway. This suggests that attributing stronger agency to the system may independently intensify oppositional reactions. Thus, belief updating did not merely alter surface evaluations. In contrast it restructured the cognitive representation of the interaction along two dimensions: perceived loss of personal control and increased attribution of system agency.

In summary, these results suggest that awareness interventions can recalibrate autonomy judgments and motivational responses, particularly when restrictions were previously opaque. However, the asymmetry between nudging and covert restriction highlights that awareness operates differently depending on the initial interpretative frame. Whereas overt nudges may trigger immediate autonomy concerns despite preserving freedom of choice, covert restrictions may remain below the threshold of perceived autonomy threat until they are explicitly disclosed to users. In line with prior transparency research (Susser et al., 2019), awareness alone is not universally effective; rather, its impact depends on how it reshapes users' autonomy appraisals and agency attributions within a given interaction context.

While the current section describes how belief updating partially recalibrates autonomy perceptions, the question remains through which psychological mechanisms these changes translate into reactance responses. We therefore turn to the underlying processes linking autonomy judgments, agency attribution, and motivational reactions in the next section.

4.3. Psychological mechanisms underlying technology reactance

Beyond demonstrating condition differences and temporal shifts in autonomy perceptions, the present findings provide insight into the psychological mechanisms through which technology paternalism translates into motivational reactance. Against the background of PRT (Brehm, 1966; Miron & Brehm, 2006), we assumed that reactance emerges primarily when individuals perceive threats to their behavioral freedom rather than merely when objective restrictions exist. The results support this interpretation and highlight that the relationship between paternalistic system behavior and reactance is mediated by users' interpretations of autonomy and agency.

Perceived autonomy as a central mechanism. Consistent with PRT, mediation analyses revealed that changes in perceived personal autonomy constituted the strongest pathway linking paternalistic conditions to reactance responses (Autonomy-Loss Pathway). Reductions in perceived personal autonomy predicted increases in reactance, suggesting that motivational resistance emerges primarily when users interpret system behavior as limiting their own agency. This finding aligns with prior work emphasizing that autonomy threat appraisal is a central antecedent of reactance (Miron & Brehm, 2006). Importantly, this mechanism operated most strongly in restrictive conditions, particularly when participants became aware of previously hidden constraints.

System agency attribution as a complementary pathway. In addition to personal autonomy loss, increases in perceived system autonomy contributed to reactance responses, although to a lesser degree (Agency-Attribution Pathway). This suggests a second psychological pathway in which users attribute greater agency or control to the AI system itself which intensifies perceptions of external interference. Previous research has highlighted that perceptions of technological agency influence how users interpret automated decision-making (Kim & Duhachek, 2020), and the present findings extend this idea by showing that agency attribution may independently contribute to reactance dynamics. Thus, reactance appears to emerge not only from perceived loss of self-determination but also from increased attribution of autonomous intent to the system.

Dual-pathway structure of reactance formation. Taken together, the results support a dual-process interpretation of technology reactance (see Fig. 7). Different degrees of technology paternalism influence reactance through two partially independent mechanisms: (1) an autonomy-loss pathway, whereby individuals experience reductions in personal autonomy, and (2) an agency-attribution pathway, whereby the system is perceived as more autonomous or controlling. While the autonomy-loss pathway showed stronger effects, both mechanisms jointly contributed to the observed reactance patterns.

Belief updating as a corrective but incomplete mechanism. The belief updating manipulation provides additional insight into these processes. Consistent with prior findings on awareness interventions in AI interaction (Sankaran et al., 2021; Wachner et al., 2021), increasing transparency did not uniformly alter responses but instead selectively recalibrated autonomy perceptions. In particular, covert restrictions which were initially perceived as non-threatening elicited substantial shifts on all evaluations after disclosure, suggesting that reactance depends on cognitive interpretation rather than objective restriction alone. However, recalibration was only partial, because the nudging condition showed comparatively smaller or inconsistent changes across evaluations.

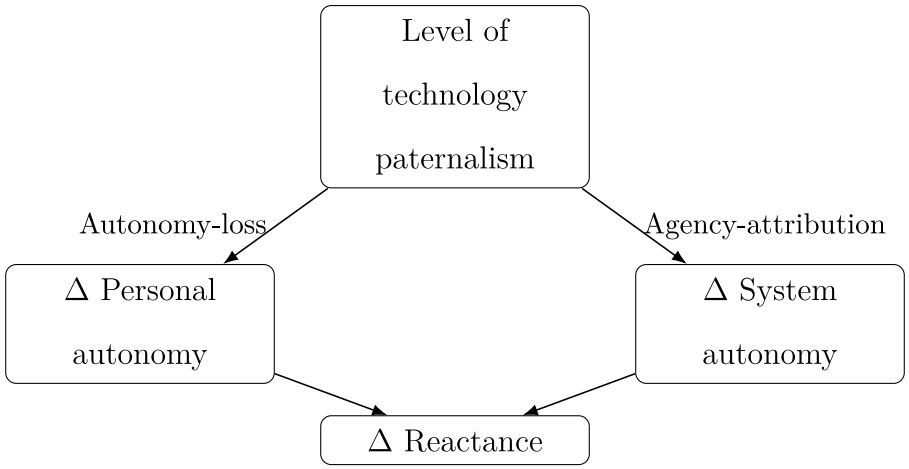


Fig. 7. Conceptual model illustrating dual pathways through which technology paternalism may influence changes in psychological reactance: via perceived losses of personal autonomy (autonomy-loss pathway) and via increased attribution of autonomy to the system (agency-attribution pathway).

Role of individual differences. We further explored whether individual traits moderated these psychological processes. Linear mixed-effects models including reactance proneness, self-efficacy, and moral agency indicated that individual differences primarily influenced initial evaluations (before belief updating) of autonomy and reactance across conditions, reflected in several Condition × Trait interactions. However, the critical Condition × Time × Trait interactions (after belief updating) were not robust, suggesting that the belief updating effect itself operated largely similarly across trait levels. From a PRT perspective, this pattern suggests that dispositional factors shape the overall sensitivity to autonomy-relevant interventions but do not fundamentally alter the underlying mechanism through which awareness of restrictions recalibrates autonomy perceptions and reactance responses. Therefore, we can constitute that traits modulate how strongly individuals evaluate AI systems, however, the dynamic shift following disclosure of paternalistic mechanisms appears to represent a comparatively general psychological process.

Implications for psychological reactance theory in AI contexts. Overall, the findings suggest that classical assumptions of PRT extend to AI-mediated environments but require refinement to account for opacity and interpretative uncertainty. Whereas traditional reactance research often assumes that threats to freedom are directly observable, AI systems introduce ambiguity that allows autonomy threats to remain unnoticed until explicitly revealed. Consequently, reactance may emerge through delayed cognitive appraisal processes rather than immediate perception. The dual-pathway model proposed here highlights how both autonomy loss and system agency attribution jointly shape users' motivational responses to paternalistic AI interventions.

5. Limitations and future work

One limitation of this work is that it was conducted online, which presumably resulted in lower immersion into the scenario than one could expect in a controlled laboratory setting. While conducting the study online increases ecological validity, because participants interacted with the prototypes in environments similar to those in which they would normally browse e-commerce platforms, it also implies that participants may have been distracted by their surroundings. Such distractions are not trivial. As the present findings suggest that autonomy perceptions depend strongly on awareness and interpretation processes, divided attention may directly affect whether paternalistic system behavior is correctly recognized. Future studies should therefore systematically contrast distracted and non-distracted interaction contexts to examine how attentional resources influence the perception of autonomy restrictions and subsequent reactance responses.

Relatedly, the visibility and salience of visual and textual cues warrant more systematic investigation. In the present study, the nudging condition elicited comparatively strong reactance reactions, which may partly be attributable to its highly salient visual design. In contrast, restrictive conditions with lower visual prominence were less likely to be immediately recognized as autonomy-relevant. Future research should experimentally vary the prominence, frequency, and modality of informational cues to better understand how initial awareness can be increased without inadvertently intensifying reactance. Combining self-report data with behavioral measures and process-tracing methods such as eye tracking would allow for more precise conclusions regarding whether and how participants attend to and process information about system influence.

Another important avenue for future work concerns the proposed dual-pathway model of technology reactance. While the present mediation analyses provide initial support for two psychological mechanisms further research is needed to test this model more directly. For example, experimental manipulations could independently vary perceived system agency and perceived restriction strength to disentangle their relative contributions. Longitudinal designs could also investigate whether repeated exposure to paternalistic AI systems strengthens or weakens these pathways over time, potentially leading to habituation or normalization effects.

Moreover, some participants indicated in open-ended responses that the AI tool was not personally relevant to them (e.g., "I couldn't see myself in the task as I manage my finances well."). This perceived lack of relevance may have influenced both autonomy judgments and reactance responses. Scenario-based designs always involve a certain degree of abstraction from real-life decision-making. Future studies should therefore explicitly measure perceived personal relevance and investigate how relevance moderates both autonomy perceptions and motivational reactions. In addition, given prior work showing that perceptions of AI autonomy differ across domains (e.g., autonomous driving vs. e-commerce; Sankaran et al., 2021), it would be valuable to replicate the present design in other AI application areas. Particularly promising contexts include news recommender systems, and social media platforms, where paternalistic influences may operate in more subtle or socially consequential ways; or in health-related decision support which is a high stakes domain.

A further limitation concerns the manipulation check. Due to ambiguous phrasing, the originally intended manipulation check could not be used in its planned form. Although participants were informed about their assigned condition during belief updating, it remains uncertain to what extent this information was attentively processed. Future studies

should employ clearer and more robust comprehension checks to ensure that participants fully understand both the system's functionality and the experimental manipulation.

Finally, the generalizability of the findings is limited by the predominantly German sample. Cultural differences in the value ascribed to autonomy and authority (Jonas et al., 2009) may influence both autonomy perceptions and reactance tendencies. Cross-cultural research would therefore be an important next step to examine whether the dual-pathway structure of technology reactance holds across cultural contexts or whether certain mechanisms are more pronounced in individualistic versus collectivistic societies.

Beyond these specific limitations, the present framework leaves room for the inclusion of additional psychological variables. For example, perceived threat to autonomy, cognitive dissonance, trust in AI systems, or prior experience with automated decision-making could potentially shape how paternalistic interventions are interpreted. Although individual differences such as reactance proneness were explored in the present study, more comprehensive modeling of dispositional and situational factors is needed to fully understand the boundary conditions of technology reactance.

Finally, future research should aim to refine the proposed dual-pathway model, systematically manipulate awareness and salience, and test the robustness of these mechanisms across domains and cultural contexts. Such work would contribute to a more nuanced understanding of how users cognitively represent and emotionally respond to paternalistic AI systems.

6. Conclusion

The present study investigated how different forms of AI-based paternalism shape autonomy perceptions and psychological reactance, with a particular focus on whether users exhibit a potential reactance deficit when restrictions remain unnoticed. Across conceptual, theoretical, and empirical levels, this work contributes to a more nuanced understanding of technology paternalism in human-AI interaction.

On a conceptual level, we contribute a refined definition of technology paternalism that differentiates between ontological aspects of system restriction and epistemic aspects related to user awareness and interpretation. This distinction allows for a more precise analysis of when paternalistic interventions become psychologically relevant to users.

On a theoretical level, we introduce a framework linking paternalistic system behavior, perceived autonomy, belief updating, and psychological reactance. Grounded in PRT, our findings suggest that reactance depends less on objective restrictions alone than on how such restrictions are perceived and interpreted. The proposed dual-pathway model highlights that changes in perceived personal autonomy and system agency represent distinct psychological mechanisms through which paternalistic interventions influence user responses.

On an empirical level, we provide experimental evidence comparing overt and covert forms of AI paternalism. The results indicate that covert restrictions may initially remain unnoticed and therefore fail to activate protective motivational responses, whereas visible interventions such as nudging can evoke reactance even when objective restrictions are limited. Belief updating partially corrected miscalibrated autonomy perceptions, demonstrating that awareness processes play a crucial role but do not fully eliminate discrepancies between perceived and actual paternalism.

In summary, the findings underline that decision-making processes supported by AI systems can become increasingly opaque, potentially weakening users' ability to recognize autonomy-relevant interventions. By integrating conceptual clarification, theoretical development, and experimental evidence, this work advances the emerging research area of technology paternalism and provides a foundation for future studies examining how AI systems can be designed in ways that preserve human autonomy while supporting decision-making.

CRediT authorship contribution statement

Michelle Görlitz: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Astrid Marieke Rosenthal-von der Pütten:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Formal analysis, Conceptualization.

Publication ethics

Informed consent was obtained from all participants included in the study. All procedures in studies involving human participants were performed in accordance with international standards. The Independent Ethics Committee of the Faculty of Medicine at RWTH Aachen University provided ethical clearance.

Declaration of Generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used ChatGPT 5.2 in order to (1) revise language (orthography and grammar checks), (2) prepare tables in latex, (3) resolve compiling errors when writing in latex and to resolve compiling errors in R code. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This research was not supported by any grant.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Full introductory text

In this study, a new AI budgeting system is being tested. Before you start interacting with the prototype, a short introduction with information about the system and the scenario you are about to find yourself in:

Imagine this for the rest of this study: You are struggling with your personal finances and have decided that you need help making better purchasing decisions. You have heard about the new AI Financial Assistant and have decided to download and use the system.

What can the AI financial assistant do?

- Monthly invoice tracking - Payment reminders - Expense tracker - Assistance for better purchasing decisions - And much more

How does the AI financial assistant work?

The AI-supported system analyzes your personal finances by accessing your banking and payment apps, identifies patterns and derives specific monthly budgets for individual areas of your life to help you get your finances under control.

Scenario: Imagine interacting with the prototype you want to buy a new item of clothing (more detailed instructions follow) and visit an online shop. You want to start your search. Before you do so, however, a pop-up window appears informing you that the download of the AI assistant is complete.

If the AI financial assistant is active, you can recognize this by the green logo (see below) at the top right of the browser window.

Data availability

RawDataPaternalismStudy2024Anonymized.xlsx (Original data) (Open Science Framework)

References

Ahmad Husairi, M., & Rossi, P. (2024). Delegation of purchasing tasks to AI: The role of perceived choice and decision autonomy. *Decision Support Systems*, *179*, Article 114166. http://dx.doi.org/10.1016/j.dss.2023.114166.

APA Dictionary of Psychology (2018). Paternalism.

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611–623. http://dx.doi.org/10.1007/s00146-019-00931-w.

Baird, A., & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, *45*(1), 315–341. http://dx.doi.org/10.25300/MISQ/2021/15882.

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, *4*(3), 359–373. http://dx.doi.org/10.1521/jscp.1986.4.3.359.

Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, *90*, 363–371. http://dx.doi.org/10.1016/j.chb.2018.08.028.

Barton, A., & Grüne-Yanoff, T. (2015). From libertarian paternalism to nudging—and beyond. *Review of Philosophy and Psychology*, *6*, 341–359. http://dx.doi.org/10.1007/s13164-015-0268-x.

Brauner, P., Hick, A., Philipsen, R., & Ziefle, M. (2023). What does the public think about artificial intelligence?—A criticality map to understand bias in the public perception of AI. *Frontiers in Computer Science*, *5*, http://dx.doi.org/10.3389/fcomp.2023.1113903.

Brehm, J. W. (1966). A theory of psychological reactance. *Oxford, England: Academic Press*.

Bruns, H., & Perino, G. (2023). The role of autonomy and reactance for nudging — Experimentally comparing defaults to recommendations and mandates. *Journal of Behavioral and Experimental Economics (Formerly the Journal of Socio-Economics)*, *106*(C), http://dx.doi.org/10.1016/j.socec.2023.102047.

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. http://dx.doi.org/10.1002/bdm.2155.

Candrian, C., & Scherer, A. (2023). Reactance to Human versus Artificial Intelligence: Why Positive and Negative Information from Human and Artificial Agents Leads to Different Responses. http://dx.doi.org/10.2139/ssrn.4397618.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior* (1st ed.). Springer New York, NY.

Dillard, J. P., & Shen, L. (2005). On the Nature of Reactance and its Role in Persuasive Health Communication. *Communication Monographs*, *72*(2), 144–168. http://dx.doi.org/10.1080/03637750500111815.

Dolgopolova, I., Li, B., & Roosen, J. (2020). Do Nudges Matter? Consumer Perception and Acceptance of Recommender Systems with Different Types of Nudges. In *3rd fAccTRec workshop: responsible recommendation at the ACM conference series on recommender systems RECSYS 2020*.

Dworkin, G. (2020). Paternalism. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Fall 2020)*. Metaphysics Research Lab, Stanford University, https://plato.stanford.edu/archives/fall2020/entries/paternalism/.

Ehrenbrink, P., & Möller, S. (2018). Development of a reactance scale for human–computer interaction. *Quality and User Experience*, *3*(1), 2. http://dx.doi.org/10.1007/s41233-018-0016-y.

Eiband, M., Völkel, S. T., Buschek, D., Cook, S., & Hussmann, H. (2019). When people and algorithms meet: User-reported problems in intelligent everyday applications. In *Proceedings of the 24th international conference on intelligent user interfaces* (pp. 96–106). http://dx.doi.org/10.1145/3301275.3302262.

Fabbri, M. (2023). Self-determination through explanation: an ethical perspective on the implementation of the transparency requirements for recommender systems set by the Digital Services Act of the European Union. In *Proceedings of the 2023 AAAI/ACM conference on AI, ethics, and society* (pp. 653–661). http://dx.doi.org/10.1145/3600211.3604717.

Faraoni, S. (2023). Persuasive technology and computational manipulation: Hypernudging out of mental self-determination. *Frontiers in Artificial Intelligence*, *6*, Article 1216340. http://dx.doi.org/10.3389/frai.2023.1216340.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. http://dx.doi.org/10.3758/BF03193146.

Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: public attitudes towards overt and covert nudges. *Judgment and Decision Making*, *8*(3), 202–213. http://dx.doi.org/10.1017/S1930297500005933.

Festinger, L., Riecken, H. W., & Schachter, S. (1956). *When Prophecy Fails*. MN: University of Minnesota Press, http://dx.doi.org/10.1037/10030-000.

Fogg, B. J. (1998). Persuasive computers: perspectives and research directions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 225–232). http://dx.doi.org/10.1145/274644.274677.

GmbH, F. (2024). Figma. https://www.figma.com/.

Gonçalves, A. R., Pinto, D. C., Shuqair, S., Dalmoro, M., & Mattila, A. S. (2024). Artificial intelligence vs. autonomous decision-making in streaming platforms: A mixed-method approach. *International Journal of Information Management*, *76*, Article 102748. http://dx.doi.org/10.1016/j.ijinfomgt.2023.102748.

Hanna, J. (2018). Hard and soft paternalism. In *The routledge handbook of the philosophy of paternalism* (pp. 24–34). Routledge.

Heilman, M. E. (1976). Oppositional behavior as a function of influence attempt intensity and retaliation threat.. *Journal of Personality and Social Psychology*, *33*(5), 574. http://dx.doi.org/10.1037/0022-3514.33.5.574.

Hofmann, B. (2003). Technological paternalism: On how medicine has reformed ethics and how technology can refine moral theory. *Science and Engineering Ethics*, *9*(3), 343–352. http://dx.doi.org/10.1007/s11948-003-0031-z.

Hong, S.-M., & Faedda, S. (1996). Refinement of the Hong Psychological Reactance Scale. *Educational and Psychological Measurement*, *56*(1), 173–182. http://dx.doi.org/10.1177/0013164496056001014.

Hughes, L., Dwivedi, Y. K., Malik, T., Shawosh, M., Albashrawi, M. A., Jeon, I., Dutot, V., Appanderanda, M., Crick, T., De', R., et al. (2025). AI agents and agentic systems: A multi-expert analysis. *Journal of Computer Information Systems*, *65*(4), 489–517. http://dx.doi.org/10.1080/08874417.2025.2483832.

Jonas, E., Graupmann, V., Kayser, D. N., Zanna, M., Traut-Mattausch, E., & Frey, D. (2009). Culture, self, and the emergence of reactance: Is there a "universal" freedom? *Journal of Experimental Social Psychology*, *45*(5), 1068–1080. http://dx.doi.org/10.1016/j.jesp.2009.06.005.

Kachanoff, F. J., Taylor, D. M., Caouette, J., Khullar, T. H., & Wohl, M. J. A. (2019). The chains on all my people are the chains on me: Restrictions to collective autonomy undermine the personal autonomy and psychological well-being of group members. *Journal of Personality and Social Psychology*, *116*(1), 141–165. http://dx.doi.org/10.1037/pspp0000177.

Kim, T. W., & Duhachek, A. (2020). Artificial intelligence and persuasion: A construal-level account. *Psychological Science*, *31*(4), 363–380. http://dx.doi.org/10.1177/0956797620904985.

Lee, G., & Lee, W. J. (2009). Psychological reactance to online recommendation services. *Information & Management*, *46*(8), 448–452. http://dx.doi.org/10.1016/j.im.2009.07.005.

Leiner, D. J. (2024). SoSci Survey. https://www.soscisurvey.de.

Mahmud, H., Islam, A. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, Article 121390. http://dx.doi.org/10.1016/j.techfore.2021.121390.

Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(2), http://dx.doi.org/10.17169/fqs-1.2.1089.

Maze Design Inc. (2024). Maze. https://app.maze.co/.

Michaelsen, P., Johansson, L.-O., & Hedesström, M. (2021). Experiencing default nudges: autonomy, manipulation, and choice-satisfaction as judged by people themselves. *Behavioural Public Policy*, *8*(1), 85–106. http://dx.doi.org/10.1017/bpp.2021.5.

Mihale-Wilson, C. A. (2025). Invisible Paternalism in Agentic IS: Rethinking Autonomy in the Design of AI-based Support.

Milano, S., Mittelstadt, B., Wachter, S., & Russell, C. (2021). Epistemic fragmentation poses a threat to the governance of online targeting. *Nature Machine Intelligence*, *3*(6), 466–472. http://dx.doi.org/10.1038/s42256-021-00358-3.

Miron, A. M., & Brehm, J. W. (2006). Reactance Theory - 40 Years Later. *Zeitschrift Für Sozialpsychologie*, *37*(1), 9–18. http://dx.doi.org/10.1024/0044-3514.37.1.9.

Oinas-Kukkonen, H., & Harjumaa, M. (2008). A systematic framework for designing and evaluating persuasive systems. In *International conference on persuasive technology* (pp. 164–176). Springer, http://dx.doi.org/10.1007/978-3-540-68504-3_15.

Olari, V., & Romeike, R. (2021). Addressing AI and Data Literacy in Teacher Education: A Review of Existing Educational Frameworks. In *Proceedings of the 16th workshop in primary and secondary computing education* (pp. 1–2). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3481312.3481351.

Oviatt, S. (2021). Technology as Infrastructure for Dehumanization: Three Hundred Million People with the Same Face. In *Proceedings of the 2021 international conference on multimodal interaction* (pp. 278–287). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3462244.3482855.

Prunkl, C. (2022). Human Autonomy in the Age of Artificial Intelligence. *Nature Machine Intelligence*, *4*(2), 99–101. http://dx.doi.org/10.1038/s42256-022-00449-9.

Rosenthal-von der Pütten, A., & Bock, N. (2018). Development and Validation of the Self-Efficacy in Human-Robot-Interaction Scale (SE-HRI). *J. Hum.-Robot Interact.*, *7*(3), 21:1–21:30. http://dx.doi.org/10.1145/3139352.

Quick, B. L., Scott, A. M., & Ledbetter, A. M. (2011). A close examination of trait reactance and issue involvement as moderators of psychological reactance theory. *Journal of Health Communication*, *16*(6), 660–679. http://dx.doi.org/10.1080/10810730.2011.551989.

Quick, B. L., & Stephenson, M. T. (2008). Examining the Role of Trait Reactance and Sensation Seeking on Perceived Threat, State Reactance, and Reactance Restoration. *Human Communication Research*, *34*(3), 448–476. http://dx.doi.org/10.1111/j.1468-2958.2008.00328.x.

Rochi, M. (2023). Technology paternalism and smart products: Review, synthesis, and research agenda. *Technological Forecasting and Social Change*, *192*, Article 122557. http://dx.doi.org/10.1016/j.techfore.2023.122557.

Rochi, M., Rauschnabel, P. A., Renner, K.-H., & Ivens, B. S. (2024). Technology paternalism: Development and validation of a measurement scale. *Psychology & Marketing*, *41*(5), 1172–1188. http://dx.doi.org/10.1002/mar.21971.

Rosenberg, B. D., & Siegel, J. T. (2018a). A 50-year review of psychological reactance theory: Do not read this article.. *Motivation Science*, *4*(4), 281. http://dx.doi.org/10.1037/mot0000091.

Rosenberg, B. D., & Siegel, J. T. (2018b). A 50-year review of psychological reactance theory: Do not read this article. *Motivation Science*, *4*(4), 281–300. http://dx.doi.org/10.1037/mot0000091.

Sankaran, S., & Markopoulos, P. (2021). " It's like a puppet master": User Perceptions of Personal Autonomy when Interacting with Intelligent Technologies. In *Proceedings of the 29th ACM conference on user modeling, adaptation and personalization* (pp. 108–118). http://dx.doi.org/10.1145/3450613.3456820.

Sankaran, S., Zhang, C., Aarts, H., & Markopoulos, P. (2021). Exploring Peoples' Perception of Autonomy and Reactance in Everyday AI Interactions. *Frontiers in Psychology*, *12*, http://dx.doi.org/10.3389/fpsyg.2021.713074.

Satow, L. (2020). B5T®. Big-Five-Persönlichkeitstest [verfahrensdokumentation, testdokumentation und fragebogen deutsch, englisch, französisch, italienisch, spanisch]. *Leibniz-Institut Für Psychologie (ZPID) (Hrsg.)*, https://www.psycharchives.org/en/item/0330df07-144c-43b8-bb3e-ff1f5ad52f94.

Sheldon, K. M., Elliot, A. J., Kim, Y., & Kasser, T. (2001). What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology*, *80*(2), 325–339. http://dx.doi.org/10.1037/0022-3514.80.2.325.

Sommer, J., Musolino, J., & Hemmer, P. (2024). Updating, evidence evaluation, and operator availability: A theoretical framework for understanding belief. *Psychological Review*, *131*(2), 373–401. http://dx.doi.org/10.1037/rev0000444.

Spiekermann, S., & Pallas, F. (2006). Technology paternalism – wider implications of ubiquitous computing. *Poiesis & Praxis*, *4*(1), 6–18. http://dx.doi.org/10.1007/s10202-005-0010-3.

Storms, E., Alvarado, O., & Monteiro-Krebs, L. (2022). 'Transparency is Meant for Control'and Vice Versa: Learning from Co-designing and Evaluating Algorithmic News Recommenders. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–24. http://dx.doi.org/10.1145/3555130.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. *Internet Policy Review*, *8*(2), https://policyreview.info/articles/analysis/technology-autonomy-and-manipulation.

Thaler, R. H., & Sunstein, C. R. (2021). *Nudge: The final edition*. Yale University Press.

Vansteenkiste, M., Ryan, R. M., & Soenens, B. (2020a). Basic psychological need theory: Advancements, critical themes, and future directions. *Motivation and Emotion*, *44*(1), 1–31. http://dx.doi.org/10.1007/s11031-019-09818-1.

Vansteenkiste, M., Ryan, R. M., & Soenens, B. (2020b). Basic psychological need theory: Advancements, critical themes, and future directions. *Motivation and Emotion*, *44*(1), 1–31. http://dx.doi.org/10.1007/s11031-019-09818-1.

Wachner, J., Adriaanse, M. A., & De Ridder, D. T. (2020). And how would that make you feel? How people expect nudges to influence their sense of autonomy. *Frontiers in Psychology*, *11*, Article 607894. http://dx.doi.org/10.3389/fpsyg.2020.607894.

Wachner, J., Adriaanse, M., & De Ridder, D. (2021). The influence of nudge transparency on the experience of autonomy. *Comprehensive Results in Social Psychology*, *5*(1–3), 49–63. http://dx.doi.org/10.1080/23743603.2020.1808782.

Wertenbroch, K., Schrift, R. Y., Alba, J. W., Barasch, A., Bhattacharjee, A., Giesler, M., Knobe, J., Lehmann, D. R., Matz, S., Nave, G., Parker, J. R., Puntoni, S., Zheng, Y., & Zwebner, Y. (2020). Autonomy in consumer choice. *Marketing Letters*, *31*(4), 429–439. http://dx.doi.org/10.1007/s11002-020-09521-z.

Yablonskaya, T. (2020). Fashion Ecommerce Business: Best Practices and Examples. https://www.scnsoft.com/ecommerce/fashion-ecommerce. (Last accessed 03 May 2024).