URBAN DESIGN International
https://doi.org/10.1057/s41289-023-00231-2

ORIGINAL ARTICLE

Behavioral designs defined: how to understand and why it is important to differentiate between "defensive," "hostile," "disciplinary", and other designs in the urban landscape

Karl de Fine Licht[1]

Accepted: 26 August 2023
© The Author(s) 2023

Abstract
In recent years, a growing discussion about how we should design our cities has emerged, particularly for the more controversial modes of design such as "defensive," "hostile," or "disciplinary" architecture (i.e., benches on which one cannot sleep, or metal studs on which one cannot skate). Although this debate is relatively mature, many studies have argued that these design notions are undertheorized and are, thus, challenging to study from an empirical and normative perspective. In this paper, I will define the most common terms used in the literature and show how they are interconnected by utilizing a set of "conditions of adequacy" from philosophy to facilitate a more transparent and well-grounded discussion of them. Terms such as "hostile" and "defensive" design are underlined by several different phenomena, not just one as is sometimes commonly assumed. I will also show that these phenomena and their conceptualizations require—and sometimes force us to use—different moral reasons when justifying the utilization of different designs.

Keywords Definitions · Defensive design · Hostile design · Defensive architecture

Introduction

Recently, a discussion has emerged about the best ways to design our cities, engaging academics, practitioners, and the public. For example, many researchers have examined "hostile," "defensive," or "excluding" architecture or design (de Fine Licht 2017; Johnsen et al. 2021; Petty 2016; Rosenberger 2020), which aim to prevent certain behaviors in public or semi-public space. Examples of these designs include benches that deter sleeping and spikes on the ground that prevent standing. These designs, to some extent, target vulnerable groups, such as the homeless or the mentally ill (Smith and Walters 2018) who already are subject to stigmatization, and, as such, the use of hostile design has been criticized by researchers as well as the general public (de Fine Licht 2017; Johnsen et al. 2021; Rosenberger 2020).

Although the question about how we should design our cities is extensive, it has been argued that the discussion of hostile and other designs is undertheorized and that the morality regarding them is perhaps not as straightforward as it seems (Johnsen et al. 2021; Rosenberger 2020). According to these authors, we need to understand the phenomena of "hostile design" better, and we also need to examine the moral underpinnings more closely. A significant challenge in the current debate is the absence of clear and comprehensive definitions, especially for terms like "defensive design," "hostile design," and so forth. Admittedly, a plethora of different *terms* such as "excluding," "hostile," "unpleasant," and "disciplinary"—or "including" and "pleasant"—exist to describe designs. There is, however, no agreement concerning which terms should require a more precise definition or how these terms relate to each other. They are often just described in a sentence or two, which is insufficient to elucidate the underlying concept. Thus, the field is in much need of conceptual clarification (Rosenberger 2020).

In this paper, I present a set of definitions to facilitate empirical examination and moral evaluation of phenomena like hostile design. I differentiate between "hostile," "defensive," "disciplinary," "excluding," and "unpleasant" designs and discuss the complex relationship between terms and how

✉ Karl de Fine Licht
Karl.definelicht@chalmers.se

[1] Chalmers University of Technology Science, Technology and Society, Vera Sandbergs Allé 8, 41 133 Gothenburg, Sweden

Published online: 30 October 2023

Table 1 Key desiderata for formulating a robust definition

Desiderata	Reasons for revising *x*
The coherence condition	*x* gives incoherent results
The precision condition	*x* cannot be determined in principle in categorical or in dimensional terms
The reliability condition	*x* is not easily determined in practice in observational, or in other terms
The measurability condition	*x* is not ordinally comparable in an intrapersonal- or interpersonal sense
The simplicity conditions	*x* is not disjunction of different criteria, and the concept core of *x* is not identified
The ordinary language condition	*x* is no tconsistent with—or gives an explanation to—ordinary language use
The value condition	*x* fails to capture the values ordinarily associated with *x*
The condition of normative adequacy	*x* does not help us to promote the relevant value we aim to achieve

they relate to and complement each other. I will also discuss—and to some extent produce—the "positive" counterparts of excluding or hostile designs: "friendly," "offensive," "accommodating," "including," and "pleasant" designs. These definitions draw from a set of well-established criteria in moral philosophy for robust definitions (see e.g., Brülde 2000 and de Fine Licht and Brülde 2021; for defense, see de Fine Licht and Folland 2019). Defining these terms with this approach will ensure that the discussion about how we should define these concepts (and perhaps the concepts in general in urban development) is transparent and well grounded, and, in doing so, the definitions themselves will, I hope, prove both practical and validated.

The structure of this paper is as follows: First, I describe the conditions of adequacy for definitions of design. Subsequently, I discuss the current definitions of, e.g., "hostile design" and why they do not live up to the conditions of adequacy outlined below. Then I will give a new set of definitions that better fulfill these conditions, and finally, I conclude the paper and provide suggestions on future research directions.

Conditions of adequacy

Conditions of adequacy are often tacitly assumed instead of explicitly stated when defining concepts. Regrettably, even though many might concur with the conditions presented here, numerous debates often employ only a subset of these conditions. This selective usage has negatively impacted the quality and applicability of the definitions (see de Fine Licht and Brülde 2021, for example). It might be that the definitions do not adhere to ordinary language, and this will result in misunderstandings of different sorts in different contexts, or it might adhere to this condition but then it is too complex, and so on.

There exists a broad spectrum of purposes and ends in defining our concepts, and various conditions may be pertinent depending on those specific purposes and ends. The goals of formulating definitions for defensive design are principally practical, yet they also bear theoretical significance. These definitions are intended for the evaluation of places, objects, and other factors in determining, for example, the types of design that should be incorporated into urban planning and development. Concurrently, there are theoretical applications for these definitions; researchers, for instance, may utilize them to enhance their understanding of how and why designs shape place and space in particular ways.

Assuming that we have these aims and try to achieve a form of broadly democratic processes and outcomes using these definitions, I have argued elsewhere that the definitions at least need to fulfill eight conditions of adequacy (de Fine Licht and Folland 2019; see Table 1 below, where *x* is the definition).

First, the definition(s) should be clear and precise (the precision condition), coherent (the coherence condition), and as simple as possible (the simplicity condition). Thus, we should have a definition that allows us to determine what, in principle, should count as defensive design instead of other forms of design. Additionally, the definition should be coherent because it does not mark a design as both defensive and not defensive at the same time. It should also be as simple as possible, thus, capturing the core concepts. Next, the definition should be operational (the reliability condition), so we can use it to determine whether something is a defensive design in practice. It is also beneficial if it can explain why, e.g., hostility can come in degrees (the measurability condition).

Second, the definition should also abide by ordinary language (the ordinary language condition) and value (the value condition), and it should be normatively adequate in the sense that it should not undermine the very thing it aims to produce (the normative adequacy condition).

When it comes to ordinary language, the definition should preferably correlate with or explain the everyday language use by both the users and the suppliers, i.e., urban planners, designers, etc. In other words, it should be able to provide an account of why we all call a range of different objects "defensive" and so on. The value condition is,

strictly speaking, a part of the ordinary language condition, but it makes us take notice of the values we usually connect with "hostile" or "friendly" designs and makes us remember why we deem these designs important in the first place. This condition limits the tendency to remove values to arrive at a simpler definition that is easier to use in practice. Of course, sometimes the conditions clash, and then we must choose which one takes precedence. However, this is problematic; we should be open with our choices (and why we have selected them) clearly for everyone to understand, not sweep them under the proverbial rug.

We should aim to fulfill these desiderata for many reasons, but the most generally applicable one is the democratic credo of trying to produce an open discourse where most people can join in. If some or all of these desiderata are violated, this discourse might break down or become impossible. For example, if we diverge significantly from the ordinary language condition and radically change the content of the terms we use, it becomes challenging for outsiders unfamiliar with these shifts to engage in democratic discussions on pivotal policy matters that concern them. Similarly, if a term is too imprecise or overly complicated, this will also be a hindrance when discussing these policy matters. (For a description of the desiderata and arguments for them, see, e.g., Brülde 2003 or de Fine Licht and Folland 2019.)

We also want the *framework* of definitions—definitions that cluster together and aim to describe adjacent phenomena—to be as simple as possible, knowing how the definitions are related while they fulfill our practical and theoretical needs. Thus, if "defensive," "hostile," and "unpleasant" designs differ in kind, we want to know precisely how they relate. Perhaps they both aim at action or behavior but are produced with different intents. We should also assess what conceptual work they do and if we, for example, can exclude some or several terms such that the overarching framework is easier to use.

Current definitions

Design, including defensive design, can refer to many things, such as physical objects, environment, space, and whole cities. These are also related in various complicated ways. To mention two counterintuitive examples, we can use defensively designed objects to produce friendly environments and defensive space to produce friendly cities (de Fine Licht 2020).

There are few thoroughly developed definitions of "defensive design" in the debate. However, some definitions capture the phenomena to some extent, and these are of great use when developing a comprehensive account. One of the most cited definitions of hostile design is the following:

Picture 1 A former bench at a playground in Gothenburg, Sweden, where people used to sit, drink alcohol, and use drugs

> hostile design […] refers to objects within public spaces that have the effect of targeting vulnerable groups, and which have garnered criticism (or should be criticized) for this hostility. (Rosenberger 2020, p. 884)

Although this definition is quite simple, straightforward, precise, and intuitive in many ways, it still has some problems. First, it is unclear what is meant by "object" here. There are at least a few design measures we plausibly would want to include in the definition of hostile design, but they are not objects in the conventional sense of physical things. For example, a common measure used to deter people from sitting in public spaces and consuming alcohol or drugs is to remove existing benches—such as what occurred in the depicted location. (See Picture 1.) Similarly, when designing one of the largest public transport hubs in Sweden, Brunnsparken, the planners did not include a roof because they believed it might attract the "wrong people."[1] Designing away certain groups or behaviors is clearly a part of the phenomena we are interested in when discussing hostile design, even if a physical object is not included, and we need to make this clear in the definition.

Second, it is somewhat unclear whether the intent of the designer, city planner, or relevant actors is part of Rosenberger's definition, since, for something to be considered a design, it needs to be part of a purposeful process, implying intentionality. There are several possible intentions behind one and the same design, such as to defend, degrade, or discipline, among others. The purposes of our actions and how

[1] This came up in an interview I conducted with the people in charge of the development process.

Picture 2 Skate stoppers installed to prevent skateboarders from grinding on the steps in Melbourne, Australia

we approach one another have significant moral importance in most ethical discourses and theories, as exemplified by the special laws addressing hate crimes in many countries today (Brax and Munthe 2015).

Similarly, employing defensive measures when developing a public square could be perceived as more egregious if done with malicious intent toward a vulnerable group, compared to acting in "what is best" during a challenging situation. For a comprehensive analysis of policy decisions, it's essential to discern the purposes and aims of policymakers. However, the debate often conflates terms such as "hostile," "defensive," and "disciplinary" design, even though they likely represent different intentions and consequently describe distinct phenomena (refer to Table 3, Appendix A, for definitions). It is crucial to recognize that these terms do not inherently signify the same morally significant phenomena.

Furthermore, the concept of unintentional design—where an outcome is inadvertently realized as a by-product or other means during a design process—warrants acknowledgment from a moral perspective. It brings into play other moral reasoning, such as those concerning culpable ignorance. That said, intentionality should be central to the definition. This ties back to the second point about the vague distinction between design terms, underscoring the need to understand the subtle differences among them to gauge the moral ramifications of design choices more accurately.

Third, from both literature and public discourse, it's evident that we should consider skatestoppers—devices that deter or prevent skateboarding—as part of our discussions. (Refer to Picture 2.) Given this, it becomes somewhat implausible to include "vulnerable groups" within the definition of defensive design. While some skateboarders can be

Picture 3 Anti-pigeon spikes in Gothenburg, Sweden

deemed vulnerable—many are young, lacking voting rights or resources to challenge these designs—it is worth noting that they often hail from middle-class backgrounds (de Fine Licht 2017). This demographic receives significant media attention and is frequently incentivized to vacate certain public areas in exchange for designated skate parks. If we operate under the assumption that middle-class children are the primary affected group and do not consider them "vulnerable," many would still classify skatestoppers as designs targeting skateboarding. Thus, while the primary instances of defensive design might focus on vulnerable groups, it shouldn't be a requisite in defining what constitutes defensive design.

Fourth, linked to the third point, it remains ambiguous whether "groups" encompasses non-human animals. In public discussions and academic debates, defensive design typically includes non-human animals, like pigeons and rats, or polar bears in places like Alaska (refer to Savić, 2013) (See Picture 3). While we often refer to non-human animals in group terms—such as a flock of pigeons or a mischief of rats—it is not always consistent. Therefore, clarity is essential to indicate that both human and non-human animals are considered when discussing entities targeted by defensive design.

Fifth, in-depth analyses by myself (2017) and Johnsen et al. (2021) on the moral status of hostile design did not determine that all such designs are inherently morally wrong. Therefore, not all such designs should be criticized

on those grounds. By extension, the definition should not imply that the design is, or ought to be, criticized. If so, discussions by me and Johnsen et al. would not pertain to the acceptability of defensive design but to a different matter. Admittedly, one could posit that only designs with moral issues are genuinely hostile. However, this stance would lead to intricate definitions and might misalign with conventional language and values.

The most comprehensive account to date comes from myself (de Fine Licht 2020). While it marks some progress, it grapples with numerous challenges. Specifically, it tends toward complexity. The relationships among terms lack clarity, and the definitions do not adequately differentiate between a design's intent and its tangible outcomes or between design polarities (e.g., from friendly to hostile, or accommodating to disciplinary). Clearly, this field warrants further exploration. (For a summary of the definitions addressed here, refer to Table 3 in Appendix A).

New definitions

When delving into concepts like "hostile design" in the realm of urban planning and design, it seems that fundamentally, we are discussing attempts to shape specific outcomes by influencing human or non-human behaviors. Such manipulation can e.g., be realized by altering group dynamics in a given space or by excluding certain groups—like packs or cliques—from accessing that space. For the purposes of this paper, I call this "behavioral design."

Consider the following: if space is deliberately designed to prevent people from sleeping because such activity might make others feel unsafe, then we have an agent (A) using a means (M) to modify behavior (B) with the intent to produce a specific outcome (O). Behavioral designs can be targeted to affect the actions of individuals or groups both within and outside the targeted space—for instance, by prohibiting certain groups from entering a location. Classic examples include the use of barbed wire to deter unwanted individuals and animals, or car barriers to restrict unwanted vehicles.

It is also essential to recognize that behavioral design can manifest either intentionally or unintentionally. For example, a design feature like handles on benches, intended to aid those with disabilities, may inadvertently deter people from sleeping on the bench. Such unintended consequences mark an instance of unintentional behavioral design. Thus, the foundational structure of behavioral design can be summarized as:

Behavioral design : *M* should be judged as "behavioral design" when agent *A* tries or merely succeeds in preventing or realizing an outcome *O* by modifying the behavior *B* of group *G, P,* etc., with the means *M*.

There are also adjacent phenomena, such as "affective design" or "perceptual design," which can be defined in the following way:

Affective design : *M* should be judged as "affectual design" when agent *A* tries or merely succeeds in producing a change in affect in *G, P, etc.* with the means *M*.

Perceptual design: *M* should be judged as "perceptual design" when agent *A* tries or merely succeeds in producing a perceptual state in *G, P, etc.* with the means *M*.

In the context of affective design, for instance, trees might be removed to engender feelings of safety among the populace. In perceptual design, trees might be removed to increase brightness. However, this paper chiefly zeroes in on behavioral design since it appears central to the prevailing debate on defensive design.

Defensive and offensive design

The various types of behavioral design possess different moral valences, making it vital to distinguish among them. Nevertheless, a shared trait among them is that they are intentional designs. The primary categories in behavioral design encompass "hostile" and "friendly design", "defensive" and "offensive design", "including" and "excluding design", "disciplining", and what I term "accommodating design."

Let's begin with "defensive design". Generally, when we engage in defensive behavior, it is driven by a desire to prevent a specific outcome or event. This could range from erecting embankments to thwart flooding, to constructing city walls for protection against invaders, or "parking the bus" in football to prevent the opposing team from scoring a goal. Given the context of behavioral design, I propose the following definition for defensive design, which addresses the previous objections:

Defensive design: M should be judged as "defensive design" when *A* tries or merely succeeds in preventing or realizing *O* by preventing *B* of *G, P*, etc., with *M*.[2]

Consequently, *M* should be understood as a defensive design if the agent or actor (*A*) produces or uses a means (*M*) to stop a behavior (*B*) from producing or hindering an outcome (*O*). The defensiveness is, thus, focused on the behavior of groups, packs, etc., and not only on outcomes, i.e., one tries or just succeeds in producing a safe space by

[2] This definition is quite close to the wordbook definition (see https://dictionary.cambridge.org/dictionary/english/hostile-architecture) and many definitions in the debate (see Johnsen et al. 2021, Chadalavada and Sripadma 2020, Chellew 2019, Newman 1973, and Petty 2016). Petty (2016, p. 68) says, for instance, that "Defensive architecture [consists of] … structures … installed in spaces of public use in order to render them unusable in certain ways or by certain groups.

Picture 4 The "Talk Bench". The sign indicates that sitting on this bench signals an openness to socializing. Location: Gothenburg, Sweden

barring people from sleeping or driving their cars in that space.

Contrasting with "defensive design" is what I call "offensive design." While the term might sound slightly peculiar, it can be interpreted in the following manner: Consider a football team on the offense, pushing aggressively to compel their adversaries into committing errors, thereby enabling them to score a goal. Drawing inspiration from this, we can define offensive design as:

Offensive design: *M* should be judged as "offensive design" when *A* tries or merely succeeds in preventing or realizing *O* by producing *B* of *G, P*, etc., with *M*.

For instance, urban planners often introduce lengthy, comfortable benches to foster opportunities for strangers to interact and for friends to congregate and socialize. This strategy exemplifies offensive design or architecture, where the objective is to stimulate specific behaviors—in this case, facilitating interpersonal interactions—to achieve desired outcomes, such as enhancing feelings of safety in public spaces. (Refer to Picture 4.)

The outlined definitions of offensive and defensive design seem to fulfill the conditions of adequacy quite well. They are simple, relatively precise, and in line with our ordinary language and values. Furthermore, the words "defensive" and "offensive" do not, in themselves, suggest that the designs are morally obligatory or prohibitory in their use, and thus a substantial debate could be allowed concerning whether these designs could be justified to utilize and, if so, when and why, which was one of the prerequisites for a good definition of these designs outlined in 3.1.

In terms of measurability, it's plausible to ascertain the extent to which intentional defensive or offensive designs are implemented. This can likely be achieved through discussions with designers and urban planners, document analysis, and observations regarding the intended purposes of specific designs. The same approach can be applied to gauge the influence of unintentional defensive and offensive designs, perhaps by assessing the aftermath of design modifications in a particular area. This rationale extends to all other definitions provided.

Furthermore, the definitions for defensiveness and offensiveness come in degrees—the more focus or defensive or offensive intent, the more defensive the design, or the more behavior change, the more defensive design. Defensive and offensive also seem to tie into two different concepts, which have to do with how we think about an object when we are trying to achieve something with it, i.e., whether we aim to produce actions or merely try to avoid them being performed.

When evaluating defensive and offensive design from a moral point of view, we are plausibly most interested in forward-looking reasons—i.e., consequentialist (or even utilitarian) reasons—or mutualist or contractualist reasons (cf. Johnsen et al. 2021). In other words, will the design, on the whole, be beneficial, or will it cause harm? Can we agree on its implementation? Whether it will do good or harm overall has to do with who is affected, how, and to what extent, as well as alternatives to utilizing defensive or offensive design. Of course, other morally important reasons exist concerning defensive design, but the forward-looking and contractualist ones are plausibly those that come easiest to mind.

Before moving on, we should note that people might perceive a design to be, for example, intentionally defensive or offensive, even though it is not. If intention is necessary for something to be considered intentionally defensive or offensive, and no such intention exists, everyone should agree that the object in question does not exhibit intentionally defensive or offensive design.

However, perceptions are important. If most users perceive a space to embody defensive or even hostile design, this carries moral significance and must be taken seriously. Questions arise when the public perceives a space as having a defensive or offensive intentional design, even if it does not. Determining how to address this issue will likely involve the same kind of moral reasoning discussed earlier, such as evaluating the benefits of such perceptions among other considerations. Thus, we need to differentiate between "perceived" and "actual" designs and examine both to make a more comprehensive moral evaluation of a place, etc. (Compare the discussion about perceived and actual legitimacy, see, e.g., de Fine Licht and de Fine Licht 2020).

Hostile and friendly design

Hostile design will here be understood as a special form of defensive or offensive design where *A* uses *M* to change *B* in *G*, *P,* etc., but where *M* is imbued with ill will toward *B* or

G, *P,* etc. or *M* has the same effects as if it was imbued with ill will toward B, G, etc.

Typically, when we are hostile toward others, we send a clear signal that they should stay away. If they do not heed this warning, we engage in hostile behavior and attempt to harm them, either physically or psychologically. For example, a generic teenager who walks around the house snarling might escalate to screaming or insulting someone who approaches them, in an attempt to deter interaction through psychological harm. Similarly, hostile design sends a clear message not to engage in certain behaviors (e.g., a place where people might want to sleep being covered with sharp spikes). If we do not heed this warning, we risk harm.[3] In this context, hostile design can be defined as:

Hostile design: M should be judged as "hostile design" when *A* tries or merely succeeds in preventing or realizing *O* by modifying *B* of *G, P*, etc., with *M*, where *M* is a threat or a realization of harm.

However, some spaces or places inherently pose threats to certain groups, even if they are not designed at all. For instance, outer space is generally considered hostile to humans. A reasonable explanation is that humans would face fatal consequences, like freezing and suffocation if exposed to space without appropriate gear. Here, the inherent danger of space, not any intent, constitutes its hostility. Thus, even if something is not intentionally crafted or planned, it can still be hostile to us and our behaviors in specific contexts. This understanding is essential when discussing the concept of hostile space. Therefore, we define it as:

Hostile space: S should be judged as "hostile space" when *S* prevents or realizes *O* by modifying *B* of *G, P*, etc. where the means *S* is a threat or realization of harm.[4]

When we merge the concepts of hostile space and unintentional design, a prevalent phenomenon in public spaces emerges: an environment that feels hostile due to certain design elements, but without any underlying hostile intent. Often, cultural factors or subconscious biases among stakeholders, like designers and urban planners, can influence decisions that inadvertently create a hostile space. But sometimes, these outcomes arise purely from unforeseen circumstances combined with economic and other non-hostile influences. To morally evaluate such a space, it's pivotal to discern these elements, recognizing any intentions or factors involved, to comprehensively grasp the context. In the remainder of this paper, however, I will not delve into the definitions of non-design or space. Instead, readers can find these in Table 4 in Appendix A.

In discussions of "hostile design" and broader behaviors of hostility or contempt, scholars often connect such actions to backward-looking reasons—essentially, concepts of basic responsibility and desert (see Pereboom 2006; Strawson 1962; and Wallace 1994). For instance, if I have committed a moral transgression and bear basic responsibility for it, then, all other things being equal, reactive attitudes like anger, resentment, or blame could justifiably be directed at me. It might even be a moral requirement to send these attitudes my way since I otherwise am not treated with respect. Conversely, if I am innocent or not accountable for the action in moral judgment, it would be unjust, given all other factors remain the same, to subject me to these reactive emotions. Therefore, "hostile design" serves a distinct functional role when juxtaposed against "defensive design": the former beckons backward-looking moral reflection, while the latter prompts forward-looking moral contemplation to a greater extent.

The aspect of hostility could elucidate why hostile design has attracted substantial criticism in recent times. When marginalized individuals, like the homeless, are exposed to hostility despite no wrongdoings on their part, the treatment seems undeserved. Directing contempt or antagonism toward these groups—whom we might have moral obligations to support—is gravely unjust. This perspective could illuminate the widespread dissent against the implementation of hostile design. Notably, the paradigm of the blameless homeless has faced challenges from major right-wing factions, such as the "Swedish Democrats" in Sweden and figures like President Donald Trump in other nations. This might imply that people in the debate might want more extensive use of hostile design when measures against these groups are discussed. Be that as it may, implementing defensive design, where elements of hostile design are absent, possesses a more extensive moral latitude. Those subjected to purely defensive designs are not targets of contempt, blame, or other reactive sentiments; they are not deemed accountable or deserving of such reactions. To show that someone deserves a certain treatment is often considered more difficult to show and defend than just showing that with a specific measure, we might reach a specific end (see e.g., Pereboom 2006; Jeppsson 2023).

We should as above distinguish between, for example, "perceived intentional hostile design" and "hostile design proper." It is easy to interpret a design as intentionally hostile if it appears to target the behavior of, for instance, an already stigmatized group. Given the weight of public perception, in addition to the actual design intent, such perceptions generally ought to be taken seriously and addressed

[3] There is plausibly another purpose of hostile design, which is to show disrespect to people or that one loathes them, and this is true even in cases when behavior cannot be changed. Accordingly, even though one might think that the spikes will not prevent people from sleeping or hanging out in a public square, for instance, one might still want to send them the hurtful message that they are not wanted there.

[4] See Chellew (2019) and Karlsson (2022) for a similar view on hostile design.

appropriately. This is true for all other designs as well, and I will not be mentioning it further in this paper.

"Friendly design" can be characterized as a distinct subset of both defensive and offensive designs. Specifically, *M* qualifies as "friendly design" when *A* aims, or simply manages to produce or hinder *O* by employing friendliness in *M* (i.e., positive attitudes or displays of goodwill) toward certain *G*s, *P*s, etc., or specific *B*s to influence these *B*s. Accordingly, friendly design can be conceptualized as follows:

Friendly design: *M* should be judged as "friendly design" when *A* tries or merely succeeds in preventing or realizing *O* by modifying *B* of *G, P*, etc., with *M*, where *M* is a promise or a realization of benefit.

When we exhibit friendly behavior, our objective is typically to benefit individuals, either psychologically or physically. For instance, the "red rose carpet" project in Malmö, Sweden, was initiated to make women feel more welcomed and foster their social interactions in public areas (Picture 5). To this end, the city involved some women from a predominantly male neighborhood in co-designing the space, ensuring it was appealing enough for them to spend their leisure time. These women envisioned an ambiance of friendliness and therefore chose colors they found appealing, such as pink and purple. Such perceptual factors contribute to the locale's friendly nature.[5]

Friendly designs are also often discussed in terms of backward-looking reasons. Because individuals exposed to friendly designs are not harmed, the moral standards are less stringent and thereby there is less upheaval when these standards are not met in comparison with hostile design. Nevertheless, it is conceivable that displaying friendliness toward groups who already have more than their fair share might be perceived as unjust and thereby create such negative reactions. For instance, favoring the urban upper-middle classes or, perhaps, disproportionately catering to men could serve as examples.

A potential argument against the definitions provided is that they may violate the coherence condition. Consider the site in Malmö; it could be perceived as both intentionally friendly and hostile, contingent upon the group in focus. While it appears friendly toward women, it could seem hostile toward men—especially if we posit that the designers sought to deter men with specific color choices that might be perceived as unfriendly.

Now, the occurrence of seemingly contradictory evaluations arising from our definitions appears to me to be an inherent feature, not a flaw. When assessing a design from diverse perspectives, we might not be examining the same situation in the same way. Rather, we may be confronting two distinct scenarios, and our definitions should guide us to see this and help us to make more transparent decisions accordingly. In the realm of urban design, it is arguably paramount to consider for whom the city is designed. If one group is treated in a friendly way while another is treated with hostility, this disparity should be promptly highlighted and justified or addressed. The definitions above help us to do this, hence the contradictory evaluations are a good thing and not a failure on their part, even though that might have been thought so at the onset.

Picture 5 The carpet of the red rose, Rosengård, Malmö, Photographer Åsa Svensson

Disciplinary and accommodating design

"Disciplinary" and what we can call "accommodating" design should instead be understood when the design is supposed to or just produce *B* of *G, P*, etc., to achieve *O* with the help of *changing the motivational structure* of *G*, *P*, etc., through *M* and *B,* or where *M* changes the motivational structure of *G* or *P* directly.

Take, for instance, a city where parking lots are strategically placed a short distance from a shopping center. Moreover, within this shopping center, the design compels individuals to use stairs unless they have physical disabilities. If the intent behind this layout is to encourage people to engage in more physical activity and, more critically, to cultivate a desire for increased exercise, then we are looking at a disciplinary design. (Refer to Picture 6 for another example of disciplinary design, namely the stand-up bench.) Likewise, a more pervasive phenomenon is the inundation of commercials in personal spaces, aimed at molding us into more enthusiastic consumers. Therefore, when public spaces are structured to subject us to such intense advertising, they

[5] Of course, as with hostile design, friendly design could also be just a design where *A* tries to make *G* better off without trying to induce any sort of behavior. For instance, planting flowers might just be an act of making people feel better about being at a space, not an ambition to increase the number of people frequenting said space. However, this is beyond the scope of this paper.

embody a disciplinary design ethos. Thus, disciplinary design can be articulated as:

Disciplinary design: M should be judged as "disciplinary design" when *A* tries or merely succeeds in modifying the current motivational structure of *G, P,* etc., to create a new motivational structure of *G*, P**, etc., which produces *B* conducive to *O*.[6]

When utilized for G's benefit, the ethical considerations related to disciplinary design likely revolve around the issue of paternalism because of the violation of people's autonomy by force or manipulation. For example, many people regard overzealous health promotion as problematic, as it may compel individuals to engage in behaviors that are good for them, even if they do not want to. This phenomenon might be, e.g., a violation of John Stuart Mill's Liberty Principle (1998 (1859)), which roughly states that it is impermissible to force people to act against their will, assuming they have no intention of harming third parties. ("Might" is of the essence here, as both Mill's Liberty Principle and these designs can be interpreted as non-paternalistic.)

When the motive to change someone's motivational structure is to produce an outcome that is better *for others*, such as making the individual a "better consumer," there will plausibly be even more concerns related to the violation of people's autonomy. Manipulation, especially when we manipulate people for others' benefit, is often seen as worse, as we are purely using those who are manipulated as mere means for others' ends instead of for their own ends. Using people as mere means for their own ends is something to which they may have an easier time rationally consenting than being used as mere means for others' ends.[7]

Accommodating design is when a design attempts to provide G with what they currently desire, thereby creating an O where, for example, G's and P's preferences etc. are satisfied. Admittedly, accommodating design is not so much "behavior-changing" as it is focused on fulfilling people's and others' preferences. This is why I use it as a starting point in Fig. 1, Appendix B, where I list all the designs in a flowchart. Regardless, it could still be worth including as a contrary opposite to "disciplinary design." In light of this, accommodating design can be defined as follows:

Accommodating design: *M* should be judged as "accommodating design" when *A* tries or merely succeeds in preventing or realizing *O* by using *M* to satisfy the current needs (etc.) in the motivational structure of *G, P,* etc.

[6] For a similar view on disciplinary design, see Harvey (2002) and Peck and Tickell (2002). It should also be noted that not everyone uses the term "disciplinary" in the way I have here. For example, Bergamaschi et al. (2014) use the term to denote that we keep people away from a certain spot and not that we are aiming to shape their motivational structure.

[7] This is of course not always true, but it is a good rule of thumb to use in practice.

Picture 6 Stand-up bench, Norsesund, Sweden

Including and excluding design

Lastly, there is also inclusive and exclusive (inclusionary or exclusionary) design. These terms are best understood when a design aims to, or simply succeeds in, allowing or disallowing a specific group to access a particular place. The behavior in question involves a specific group entering an area or space and how they utilize it. Inclusive design seeks to accommodate various groups, while exclusive design may limit access or usage for certain groups.

The aforementioned "red rose" is a perfect example of including design. The goal was to have more women visit—and hang out—in a public space by designing a part of that space to make it happen. The planners tried to include women more in a public space, and it was thus a chief design goal. Considering this, including design can be defined as follows:

Including design: *M* should be judged as "including design" when *A* tries to prevent or realize *O* by facilitating *G, P*, etc., with *M* to *B*.

Alternatively, excluding design can be understood as attempting to hinder a particular group from being in a space. For example, panhandlers can be excluded from a square by removing benches and toilets. In this way, we are modifying the behavior of panhandlers to produce the outcome of keeping them away through the design of the

square. In light of this, excluding design can be defined as follows:

Excluding design: M should be judged as "excluding design" when *A* tries to prevent or realize *O* by hindering *G, P*, etc., with *M* to *B*.[8]

Notably, when crafting any precise definition, we must be careful in clarifying what we mean by the different variables. In particular, the specification of *M* and the connected *B* will be of significant importance when we morally evaluate the design. For instance, if we want to include women and exclude men from a specific space, this might not be that controversial, given how men and women currently utilize that space. However, should we propose that men be excluded from public space altogether, this stance would become much more contentious. To ensure a clear understanding of what is being discussed, *M* and the interrelated *B* must be defined carefully and precisely.

Including and excluding designs should plausibly be understood as a subcategory of defensive/offensive design because we aim to achieve a particular O where, for example, G is included or excluded by M through affecting, for example, B of G. These M's could be designed to be offensive or defensive, friendly or hostile, and disciplinary or accommodating.

However, it might still be valuable to retain the notions of including and excluding design. This is because targeting specific groups to be either included or excluded often carries great moral importance. For example, according to most normative theories and conventional moral thinking, we should prioritize the worst-off groups. *Prima facie*, there are minimal legitimate situations or reasons to exclude the worst-off from something beneficial, such as public space, compared with excluding the well-off. Although the other definitions might cover this priority aspect, it still seems that the terms "including" and "excluding" designs bring extra focus to the included or excluded groups, which is valuable due to the high moral significance.

Furthermore, utilizing the terms "including" and "excluding" design easily lends itself to distributive justice or the principle of equal opportunity for some goods. Including any given group can be done for various reasons; for the carpet of the red rose, was associated with giving women a greater and more equal opportunity to share the benefits of public space. Similarly, when excluding any given group, the issue should focus on the extent to which that group receives a less equal opportunity in that space.

The main problem with utilizing the terms "excluding" and "including" design is that they might lead to confusion since exclusion and inclusion can be done without, for example, any hostility or friendliness. Take, for example, a definition previously proposed:

"Defensive" or "hostile" architecture is designed to actively exclude particular categories of persons. This term is commonly used to describe architecture that discourages the homeless and itinerant. (Smith and Walters 2018, p. 2983f)

If we interpret this quote literally and apply the definitions presented in this paper (where defensive design does not necessarily include any hostile intent), the definition by Smith and Walters (2018) reveals that it is somewhat unclear what we are discussing. Are we referring to the exclusion of people because we feel disgusted by them? Or are we doing so for more pragmatic reasons, such as encouraging as many people as possible to visit a park? Assuming that these attitudes have moral significance, the moral question becomes somewhat obscured. However, since the use of terms like "excluding" and "including" design seems to provide more clarity than confusion, we should probably retain them.

Lastly, there are "pleasant" and "unpleasant" designs. These terms are encompassed within other design terms (e.g., "hostile" and "friendly" design), using those specific terms (e.g., "hostile" and "friendly") instead makes it easier to morally evaluate the design. For instance, we might employ unpleasant designs to alter people's motivation or to convey our resentment toward them. Here it is unclear what the intentions are, and hence it is more difficult to know what we are discussing. The benefits such as those of "excluding" and "including design" are also absent. Consequently, we should either omit these pleasant and unpleasant designs from our framework, as I have done here and as Cherwell and I have done before, or we could involve pleasant and unpleasant designs in discussions about the best ways to change behavior. (For an overview of the definitions, see Table 2 below, and for the connections between the definitions, see Fig. 1 in Appendix B)

Conclusion

This paper introduces a framework for understanding and categorizing various modes of behavioral design, moving beyond traditional definitions. It distinguishes between subcategories such as defensive, offensive, hostile, friendly, and disciplinary designs, including, and excluding designs, linking them to established ethical theories and approaches to create a rich theoretical landscape. This approach not only enhances the scholarly discourse on urban design but also provides practical insights to guide field practitioners.

[8] Davis (2006 (1990), Smith and Walters (2018), Carr (2020), and Peršak (2021) have attempted to capture the phenomenon I have described here, but sometimes in other terms.

The paper elucidates various behavioral design modes, which, although related, differ in moral valence. The primary categories are "defensive" and "offensive" designs, aiming to either hinder or promote specific behaviors through direct engagement or by affecting the user's motivation. Subcategories include hostile, friendly, and disciplinary designs, characterized by their respective intents. Inclusive and exclusive designs focus on particular groups, while designs may also be classified as pleasant or unpleasant modes, a distinction that may be irrelevant in this context. These different definitions let us see a more complex variety in the phenomena of "behavioral design" than previously acknowledged,

Additionally, these different definitions both clarify and sometimes necessitate distinct justifications, accentuating their practical and moral importance. For instance, defensive and offensive designs (prima facie at least) align with consequentialist and contractualist thinking, while hostile and friendly designs correlate with concepts of desert and basic responsibility. Inclusive and exclusive design relates to equality of opportunity, and disciplinary and accommodating design connects with autonomy and paternalism. Thus, this paper contributes to developing aspects of both urban design theory and practice.

When it comes to future research, key areas of focus should include the full operationalization of definitions for practical application, a more profound analysis of these definitions, and the justifications for design utilizations, as well as the application of the fully operationalized definitions to multiple urban development cases to gauge the prevalence of different design modes. This assessment might be conducted through the lens of valuation studies, for instance, an effective approach for examining value judgments and practices in urban development more broadly (see, e.g., Molnar 2023), of which defensive and other designs are part. Monitoring these changes over time is equally critical, as the design landscape may shift in conjunction with the political environment. While this paper has initiated this essential work, further exploration and analysis are necessary to deepen our understanding and application of these complex design modalities.

Table 2 Types of designs and their corresponding definitions

Type of Design	Definition
Behavioral design	M should be judged as "behavioral design" when agent A tries or merely succeeds in preventing or realizing an outcome O by modifying the behavior B of group G, P, etc., with the means M
Defensive design	M should be judged as "defensive design" when A tries or merely succeeds in preventing or realizing O by preventing B of G, P, etc., with M
Offensive design	M should be judged as "offensive design" when A tries or merely succeeds in preventing or realizing O by producing B of G, P, etc., with M
Hostile design	M should be judged as "hostile design" when A tries or merely succeeds in preventing or realizing O by modifying B of G, P, etc., with M, where M is a threat or a realization of harm
Friendly design	M should be judged as "friendly design" when A tries or merely succeeds in preventing or realizing O by modifying B of G, P, etc., with M, where M is a promise or a realization of benefit
Disciplinary design	M should be judged as "disciplinary design" when A tries or merely succeeds in modifying the current motivational structure of G, P, etc., to create a new motivational structure of G*, P*, etc., which produces B conducive to O
Accommodating design	M should be judged as "accommodating design" when A tries or merely succeeds in preventing or realizing O by using M to satisfy the current needs (etc.) in the motivational structure of G, P, etc.
Excluding design	M should be judged as "excluding design" when A tries to prevent or realize O by hindering G, P, etc., with M to B
Including design	M should be judged as "including design" when A tries to prevent or realize O by facilitating G, P, etc., with M to B

Appendix A

See Tables 3, 4.

Table 3 Definitions of Defensive, Hostile, Unpleasant, and Exclusionary Design

Term	Definition
Defensive design/architecture	"Defensive urban design, also known as hostile, unpleasant, or exclusionary architecture, is an intentional design strategy that uses elements of the built environment to guide or restrict behavior in urban space as a form of crime prevention, protection of property, or order maintenance. It often targets people who use or rely on public space more than others, like people who are homeless and youth, by restricting the [behaviors] they engage in." (Chellew 2019, p. 19) "'Defensive architecture' to make the built environment less conductive to 'undesirable' activities, and/or the surveillance and policing of targeted areas." (Johnsen et al. 2018, p. 1106) "'softer' situational measures meant to discourage or disable the homeless from engaging in their activities… so-called defensive architecture" (Peršak and Di Ronco 2018)
Hostile design/architecture	"hostile design […] refers to objects within public spaces that have the effect of targeting vulnerable groups, and which have garnered criticism (or should be criticized) for this hostility." (Rosenberger 2020, p. 884) "In my view, a 'hostile architecture' has emerged that targets everyone who does not fit in." (Lehr 2019, p. 157) definition of hostile architecture as follows: The ways in which various physical structures are constructed, or in most instances, how elements are installed onto pre-existing physical structures to control the use of space for certain functions, and/or by certain groups of people "Hostile architecture is generally understood to encompasses small scale implementations that affect large marginalized groups" (Carr 2020, p. 19) "hostile architecture, a design practice meant to prevent … "visible incivility". unsanctioned uses of urban space are deterred by altering the feel of the city (Lynch 2021, p. 102) "Hostile architecture is a tool used to control and regulate behavior, individuals, and groups in public spaces" (Karlsson 2022, p. 2)
Unpleasant design	[Unpleasant design is] "a collection of techniques and strategies in urban design where social control is an inherent property of objects and places." (Savičić and Savić 2013, p. 1)
Including/Excluding design	Exclusionary design is design that push the homelessness from prime to marginal spaces (Davis 1990) [Exclusionary design is when] The built environment is characterized by man-made physical features that make it difficult for certain individuals—often poor people and people of color—to access certain places (Schindler 2014, 1934, see also 1953)
Combinations	"Defensive urban design, also known as hostile, unpleasant, or exclusionary architecture, is an intentional design strategy that uses elements of the built environment to guide or restrict behavior in urban space as a form of crime prevention, protection of property, or order maintenance. It often targets people who use or rely on public space more than others, like people who are homeless and youth, by restricting the [behaviors] they engage in." (Chellew 2019, p. 19) "An exclusionary design, also referred to as defensive architecture, hostile architecture and hostile design, is a restrictively built environment that constrains the spaces that homeless individuals can occupy." (Grainger 2021, p. 4) "Defensive architecture (also referred to as hostile architecture) is a type of architecture which obstructs one's ability to comfortably spend time in a public setting … 'Defensive' or 'hostile' architecture is designed to actively exclude particular categories of person[s]. This term is commonly used to describe architecture that discourages the homeless and itinerant." (Smith and Walters 2018, p. 2983f) 'Hostile architecture', also known as 'defensive' or 'disciplinary' architecture, is a relatively recent term. It loosely describes various structures that are attached to or installed in spaces of public use in order to render them unusable in certain ways or by certain groups (Petty 2016, p. 68)

Table 4 Definitions of Various Spaces Including Defensive and Offensive Space

Design type	Definition
Defensive space	S should be judged as "defensive space" when S prevents or realizes O by modifying the behavior B of G, P, etc
Offensive space	S should be judged as "offensive space" when S prevents or realizes O by producing B of G, P, etc
Hostile space	S should be judged as "hostile space" when S prevents or realizes O by modifying B of G, P, etc., where the means S is a threat or realization of harm
Friendly space	S should be judged as "friendly space" when S prevents or realizes O by modifying B of G, P, etc., where S is a promise or realization of benefit
Disciplinary space	S should be judged as "disciplinary space" when S prevents or realizes O by modifying the current motivational structure of G, P, etc. to create a new motivational structure G*, P*, etc., which produces B conducive to O
Accommodating space	S should be judged as "accommodating space" when S prevents or realizes O by using S to satisfy the current needs in the motivational structure of G, P, etc
Excluding space	S should be judged as "excluding space" when S prevents or realizes O by hindering G, P, etc. to B
Including space	S should be judged as "including space" when S prevents or realizes O by accommodating for G, P, etc. to B

Appendix B

See Fig. 1.

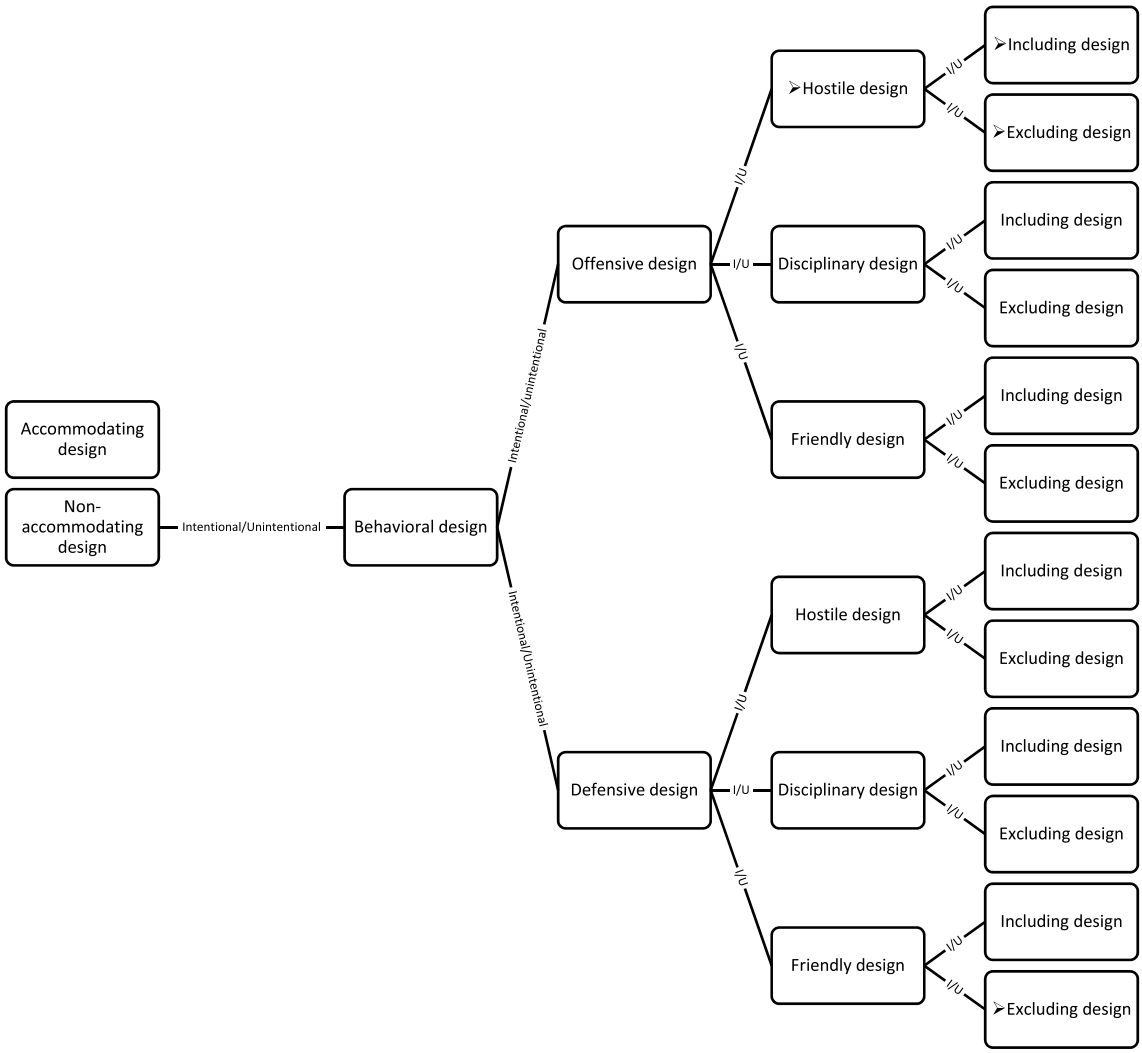

Fig. 1 Hierarchical breakdown of design approaches, distinguishing between accommodating, and non-accommodating behavioral design, and their specific sub-categories

Funding Open access funding provided by Chalmers University of Technology.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bergamaschi, M., M. Castrignanò, and P.D. Rubertis. 2014. The homeless and public space: Urban policy and exclusion in Bologna. *Revue Interventions Économiques Papers in Political Econom* 51: 1–20.

Brax, D., and C. Munthe. 2015. The philosophical aspects of hate crime and hate crime legislation: Introducing the special section on the philosophy of hate crime. *Journal of Interpersonal Violence* 30 (10): 1687–1695.

Brülde, B. 2000. On how to define the concept of health: A loose comparative approach. *Medicine, Health Care and Philosophy* 3 (3): 303–306.

Carr, M.M. 2020. Urban hostility: CPTED, hostile architecture, and the erasure of democratic public space. University Honors Thesis, Portland State University.

Chadalavada, K., and S.E. Sripadma. 2020. Defensive architecture: A design against humanity. *International Journal of Advance Research Ideas, and Innovations in Technology* 6 (1): 89.

Chellew, C. 2019. Defending suburbia: Exploring the use of defensive urban design outside of the city centre. *Canadian Journal of Urban Research* 28 (1): 19–33.

Davis, M. 1990. *City of quartz: Excavating the future in Los Angeles*. London: Verso Books.

de Fine Licht, K.P. 2017. Hostile urban architecture: A critical discussion of the seemingly offensive art of keeping people away. *Etikk I Praksis-Nordic Journal of Applied Ethics* 2: 27–44.

de Fine Licht, K., and A. Folland. 2019. Defining "social sustainability": Towards a sustainable solution to the conceptual confusion. *Etikk i Praksis-Nordic Journal of Applied Ethics* 13 (2): 21–39.

de Fine Licht, K. 2020. "Hostile architecture" and its confederates: A conceptual framework for how we should perceive our cities and the objects in them. *Canadian Journal of Urban Research* 29 (2): 1–17.

de Fine Licht, K., and B. Brülde. 2021. On defining "reliance" and "trust": Purposes, conditions of adequacy, and new definitions. *Philosophia* 49 (5): 1981–2001.

Grainger, G.L. 2021. Punishment, support, or discipline? Taking stock of recent debates about homeless governance in neoliberal cities. *Sociology Compass* 15 (8): e12909.

Harvey, D. 2002. *Spaces of capital: Towards a critical geography*. Oxfordshire: Routledge.

Jeppsson, S. 2023. Agency and responsibility: The personal and the political. *Philosophical Issues.* 5: 1–13.

Johnsen, S., S. Fitzpatrick, and B. Watts. 2018. Homelessness and social control: A typology. *Housing Studies* 33 (7): 1106–1126.

Johnsen, S., B. Watts, and S. Fitzpatrick. 2021. Rebalancing the rhetoric: A normative analysis of enforcement in street homelessness policy. *Urban Studies* 58 (2): 355–371.

Karlsson, E. 2022. Discipline and order. Master Thesis, Lund University, Lund. [online] Available at: https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9098250&fileOId=9098251

Lehr, P. 2019. *Counter-terrorism technologies: A critical assessment*. Cham: Springer.

Lynch, E. 2021. Laws of perpetual motion: The sensory regulation of mobility in public space. *The Senses and Society* 16 (1): 102–109. https://doi.org/10.1080/17458927.2020.1763037.

Mill, J.S. 1859. *On liberty and other essays*. Oxford: Oxford University Press.

Molnar, S. 2023. The framing of urban values and qualities in interorganisational settings: The case of ground floor planning in Gothenburg. *Sweden. Urban Studies* 60 (2): 292–307.

Newman, O. 1973. *Architectural design for crime prevention*. Washington, DC: National Institute of Law Enforcement and Criminal Justice.

Peck, J. and Tickell, A., 2002. Neoliberalizing space: The free economy and the penal state. In: *Spaces of Neoliberalism: Urban Restructuring in North America and West Europe*. Maiden: Blackwell Publishers Ltd.

Pereboom, D. 2006. *Living without free will*. Cambridge: Cambridge University Press.

Peršak, N., and A. Di Ronco. 2018. Urban space and the social control of incivilities: Perceptions of space influencing the regulation of anti-social behaviour. *Crime, Law and Social Change* 69 (3): 329–347.

Peršak, N. 2021. Offending sights and urban governance: Expectations of city aesthetics and spatial responses to the unsightly. In *Harm and disorder in the Urban space*, ed. N. Peršak and A. Ronco, 52–77. Oxfordshire: Routledge.

Petty, J. 2016. The London spikes controversy: Homelessness, urban securitisation and the question of 'hostile architecture.' *International Journal for Crime, Justice and Social Democracy* 5 (1): 67.

Rosenberger, R. 2020. On hostile design: Theoretical and empirical prospects. *Urban Studies* 57 (4): 883–893.

Savić, S. 2013. Unpleasant for pigeons. In: Savičić, G. and Savić, S. (eds), *Unpleasant design*. GLORIA: Kindle Version

Schindler, S.B. 2014. Architectural exclusion: Discrimination and segregation through physical design of the built environment. *Yale LJ* 124: 1934.

Smith, N., and P. Walters. 2018. Desire lines and defensive architecture in modern Urban environments. *Urban Studies* 55 (13): 2980–2995.

Strawson, P. 1962. Freedom and resentment. *Proceedings of the British Academy* 48: 187–211.

Wallace, R.J. 1994. *Responsibility and the moral sentiments*. Harvard: Harvard University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.