Let's play pretend: Towards effective modelling in experimental psycho(patho)logy

Yannick Boddez[1], Sara Scheveneels[2,3], Pieter Van Dessel[1]

1Department of Experimental Clinical and Health Psychology, Ghent University

2Department of Psychology and Educational Sciences, Open University of the Netherlands, Heerlen, The Netherlands

3Center for the Psychology of Learning and Experimental Psychopathology, KU Leuven, Leuven, Belgium

Correspondence concerning this article can be addressed to Yannick Boddez, Ghent University, Henri Dunantlaan 2 B-9000 Ghent Belgium. E-mail: Yannick.Boddez@UGent.be.

The preparation of this article was made possible by Methusalem Grant BOF16/MET_V/002 of Ghent University to Jan De Houwer and Grant FWO19/PDS/041 of the Scientific Research Foundation Flanders to Pieter Van Dessel. Sara Scheveneels is employed on KU Leuven C1 funding (C16/19/02). The first author is thankful for discussions with Bram Vervliet, Peter de Jong and Jan De Houwer about the topics of this paper.

Abstract

We introduce a novel framework to achieve more effective modelling practices in general psychology and experimental psychopathology. We first present our perspective on the meaning of the term “model”. Subsequently, we discuss three types of models that are common in psychology: laboratory models, computational models, and cognitive models. We then discuss two common ways to assess the translational value of models (phenomenological similarity and deep similarity) as well as an arguably underappreciated way (functional similarity). Functional similarity is based on an assessment of whether variables (e.g., the administration of a pharmacological substance) have a similar effect (a) in the model (e.g., in a fear conditioning procedure) and (b) on the real-life target phenomenon (e.g., on real-life anxiety complaints). We argue that the assessment of functional similarity is a powerful tool to assess the translational value of models in the field of experimental psychopathology and beyond.

Let's play pretend: Towards effective modelling in experimental psycho(patho)logy

The use of models is widespread. A fashion model may inform us how we will look with certain clothes. An architect can use a model house to gain information about how adding a wall will affect a residence. An engineer may place a small wooden car in a wind tunnel to learn about the aerodynamics of a to-be-build car (Pal, Kabir, & Talukder, 2015). In experimental psychology, deriving relations between arbitrary stimuli in an experimental task is used as a model for daily-life language (Törneke, 2010) and manipulations of the appreciation of fictitious characters is used as a model for impression formation in daily-life interactions (Gawronski, Geschke, & Banse, 2003). In experimental psychopathology, on its turn, a night of full sleep deprivation has been used as a model for insomnia (Zenses et al., 2020) and fear conditioning has been used as a model of anxiety-, trauma,- and stress-related complaints (Beckers et al., 2013; Vervliet & Boddez, 2020). The latter entails the pairing of a neutral stimulus with an aversive stimulus, which typically results in the neutral stimulus coming to elicit fear.

The use of models in science can be linked to a desire to understand, predict, and control life beyond the laboratory. This begs the question of whether these models allow us to fulfill that desire. For example, when we study fear conditioning in the laboratory, do we only learn about what happens within the limits of that procedure, or do we learn something about "real-life" psychological suffering?

In the present paper, we develop a framework for more effective modelling practices in general psychology and experimental psychopathology. We will illustrate several of our arguments by means of (fear) conditioning research, given our own background and the traction that this research has gained in the past 100 years (Vervliet & Boddez, 2020).

Model and target phenomenon

First of all, it is important to make explicit what we mean when we use the term "model" in the context of this paper. We introduce two defining criteria. First, we take the position that whether or not research relies on a modelling approach depends on whether the researcher *invokes a target phenomenon that differs from the model*. That is, a research practice only becomes a model if the researcher somehow lets it be known that the research practice (e.g., human fear conditioning in the laboratory) represents or *models* something else (i.e., the target phenomenon; e.g., real-life anxiety disorders; Weisberg, 2013). This may remind us of pretend play in children: A child's playing (e.g., with a plastic tea set) only becomes an instance of pretend play if the child indicates that they intend to *model* something else during playing (e.g., actual teatime; Deloache, 1987). Second, in research, one uses the model to *gain information* about the target phenomenon. This is what makes a model a *research model.* For example, by assessing how conditioned fear is affected by a new pharmacological treatment, a researcher may try to gain understanding of how real-life fear will be affected by this treatment.

In the section on the assessment of models below, we will pay ample attention to different ways in which a model and its target phenomenon may be similar. Still, we did not include mere similarity as a defining criterion (in fact, we made the opposite point, namely, that a model by definition differs from its target phenomenon). This is important because the same research practice can be related to an outside target phenomenon or be a target of research itself. This implies that, despite the same degree of similarity, in the first case, the research practice is used as a model, whereas in the other case it is not. For example, an operant learning task termed the Fabulous fruit game has been used as a model to gain information about addiction, but has also been treated as a study object in its own right (Buabang et al., 2021). In the first case, the Fabulous fruit game is used as a model, but in the second case it is not[1].

Modelling in psychology

In psychology, it is often a procedure that serves as model. In the context of this paper, we will term such procedures "laboratory models". Generally speaking, a procedure constitutes the basic steps executed when carrying out a study (De Houwer & Hughes, 2020). For example, in the case of human fear conditioning research, this entails presenting specific stimulus pairings (e.g., of a neutral stimulus and a fearsome stimulus) to the sampled population and assessing whether a change in fear responding occurs. Note that we consider the participant sample (e.g., a single individual like Little Albert, Watson & Rayner, 1920, or a sample of healthy first-year university students) to be part of the model (i.e., the so-called analogue sample). Just like an architect can add a miniature wall to a model house to find out how an extra wall would affect the actual residence, the researcher can play around with a laboratory procedure (e.g., administer a pharmacological substance) administered to a participant sample in order to gain information about a target phenomenon like, for example, anxiety disorders (i.e., in an ideal scenario, the pharmacological substance would affect anxiety complaints in the same way it affects conditioned fear responses). As illustrated by this example, we conceptualize the laboratory model as a starting point in research: starting from the laboratory model, one can assess how certain variables (e.g., the pharmacological substance) affect a variable of interest (e.g., fear responding), which generates information that is (either successfully or not) extrapolated to the target phenomenon (e.g., that real-life fear outside the laboratory will be affected by the pharmacological substance in the same way).

In addition to the pivotal role of procedures serving as models, psychology also makes use of computational models and of cognitive models. Although the term model is sometimes used differently from how we are using it (e.g., as synonymous for formalized theories or as theories that are narrow in scope; Fried, 2020), computational and cognitive models can also be used in the way that we propose in this paper (see Box 1 and 2).

Box 1. Computational models

Researchers who use computational models typically invoke a target phenomenon and try to gain information about that target phenomenon. Consider the Rescorla-Wagner model (1972) which gives the changes in associative strength between a CS and a US as a result of their pairings. This associative strength is assumed to map onto the strength of responding to the CS. At a computational level (Marr, 1982), the model is simply a formula that, given a certain input, specifies output. If a certain input (e.g., values for the salience of the stimuli and for the number of prior pairings) is "fed" to the model, it provides a value for the intensity or probability of responding as output. When comparing computational models to laboratory models, it is not hard to see that this is akin to what happens in a laboratory when the researcher selects stimuli of a certain salience (e.g., a tone and a shock), "feeds" these to the sampled population, and then records the intensity or probability of responding (e.g., the intensity of fear). Just like for laboratory models, a criterion for computational models to qualify as models entails that a target phenomenon (e.g., symptom levels after trauma) is invoked. This would be the case if one is not interested in "behavior" of the Rescorla-Wagner rule as such, but in how it maps onto target phenomena (which is a common scenario). For example, if one increases the value for stimulus salience, then the rule will return a higher value for responding, which can be mapped onto a more severe car accident resulting in higher conditioned fear responding. In contrast, it would not be modelling, if, for example, a mathematician simply wonders why increasing the parameter alfa in the model results in a higher output value or why the formula generates asymptotic output values without bothering about a target phenomenon.

Box 2. Cognitive models

The term cognitive model can be used in different ways (Draaisma, 2000). In the context of this paper, a criterion to speak of a model is that a target phenomenon is invoked. Suppose that the target phenomenon concerns reaction times of a given individual in a mental rotation task. In order to make predictions (e.g., to predict that identifying a full rotation takes longer than half a rotation) about this target phenomenon, one can use different types of models. For example, one could use a physical model (e.g., a little wooden windmill) or a cognitive model. From the current perspective, the unique feature of *cognitive* models is their material. Cognitive models do not exist out of wood (like the little windmill) but exist out of information processing steps. So, the material and operations of cognitive models are not physical (e.g., a rotor that puts in motion the blades of the wooden windmill), but informational in nature (Bechtel, 2008; De Houwer, Barnes-Holmes, and Barnes-Holmes, 2017; Moors, 2007). An easier way to put this would be to say that cognitive models are built from material (cognitive terms, boxes, and arrows) that is extracted from cognitive psychology textbooks.

According to the present view, cognitive models are pragmatic tools (De Houwer, 2021) that can be used to gain information about a target phenomenon in the same way as physical models, laboratory models or computational models can. The researcher walks through the information processing steps of the model (sometimes helped by boxes and arrows between them) in order to gain information about a target phenomenon (e.g., the reaction time of a given individual on a given trial in a mental rotation task). As such, these models allow researchers to develop and keep track of their inferential moves and therefore help the researcher to reason about (what will affect) the target phenomenon (i.e., they allow for surrogative reasoning; Kuorikoski & Ylikoski, 2015). Note that this view implies that the information processing steps in cognitive models do not have to be identical to the processing steps in the mind. The cognitive model is a tool that helps one to gain

information about behavior (i.e., the target phenomenon that they invoked) in the same way as, for example, a physical model would. This pragmatic stance differs from the approach in some branches in cognitive science that do not treat cognitive models as tools to predict and control behavior, but rely on behavioral experimentation with the aim to infer the actual workings of the mind (De Houwer, 2021). In such approach, both the type of statement (i.e., what something actually is rather than what something is like; Derman, 2011) and the target are different (i.e., it concerns the mind rather than behavior) from what is proposed here.

Translational value

We now turn to the evaluation of modelling practices and, in particular, to the question of how to assess the translational value of a model. Translational value refers to the extent to which knowledge obtained with the model holds for the target phenomenon. In this section as well, we start our assessment by focusing on laboratory models.

Laboratory models carry a long history of criticism. For example, Chomsky (1959, p. 30) criticized Skinner (who, for example, used operant conditioning in pigeons as a model for gambling addiction in humans), because “he uses the experimental results as evidence for the scientific character of his system of behavior, and analogic guesses ... as evidence for its scope". In current times as well, researchers question the studying of “toy problems” if the ultimate aim is to remedy real-life problems (Lewis & Wai, 2021).

A typical defense for using models in science is that some target phenomena are inaccessible (e.g., a black hole), so that studying models (e.g., a Bose–Einstein condensate) is the only option (Demirkaya, Dereli, & Güven, 2019). Relatedly, experimental psychopathologists often invoke ethical and practical reasons for studying models instead of studying the target phenomenon itself. For example, if one wants to study whether a single dose

of a pharmacological substance in the immediate aftermath of a sexual assault may serve to reduce the development of intrusions, then it may be wise to gain information in a model (see below) before trying out a new (and possibly unpromising or even harmful) intervention in an assaulted individual.

However, as hinted at in the paragraph above, this defense line for the use of models only holds if one uses models with high translational value. If not, then the information gained from the model will not translate to the target phenomenon. Perhaps surprisingly, the translational value of models in experimental psychopathology receives little attention (for exceptions see Vervliet & Raes, 2013). While it would be hard to publish a questionnaire without validation data, the same rigor is not often found in the field of experimental psychopathology (e.g., Lange, Papalini, & Vervliet, 2021). For example, fear extinction has been termed the laboratory model of choice for exposure therapy. Accordingly, a wide research community is invested in finding strategies to optimize extinction learning under the assumption that these strategies will also serve to optimize exposure therapy (Craske, Hermans, & Vervliet, 2018; Pittig, van den Berg, & Vervliet, 2016). At the same time, it is not self-evident that fear extinction research informs clinical exposure therapy and it has rarely been investigated whether the results of extinction experiments translate to exposure therapy (Scheveneels, Boddez, Vervliet, & Hermans, 2016; Scheveneels, Boddez, & Hermans, 2021). Even though most research papers in the field of experimental psychopathology remain silent about this matter, some researchers have - more or less explicitly - called attention to it. Below, we first discuss some of the often-heard arguments in the literature that focus on phenomenological similarity. Subsequently, we will focus on another form of similarity, namely functional similarity, and on how it differs from so-called deep similarity[2].

Phenomenological similarity

Phenomenological similarity refers to similarity in the way in which things appear to us via our senses without bothering about causality. At least some discussions about the translational value of models may be ingrained in the very concept of (phenomenological) similarity: Any two phenomena "share infinitely many properties and are divided by infinitely many properties" (Stanford Encyclopedia of Philosophy, 2019). So, without constraints on what counts as (sufficiently) similar, defendants and critics of a model may both feel as if they are right, as one can always point to overlap and distinctness between model and target phenomenon, respectively.

In the field of experimental psychopathology, this has led to lively discussions. For example, it has been argued that fear conditioning procedures which rely on the use of an electric shock as unconditional stimuli can at best be models for electrophobia (Wilhelm, 2021). Similarly, the use of geometrical figures as fear-conditioned stimuli has been questioned in favor of multisensory stimuli (Waters, LeBeau, & Craske, 2017), or joystick movements in studies on fear of movement (Meulders, 2020), and of pictures of gas stoves in studies on obsessive compulsive disorder (Krypotos & Engelhard, 2020). Authors have also made suggestions to make extinction experiments look more like exposure therapy sessions (Kredlow, de Voogd, & Phelps, 2020). These examples illustrate that the research community is focused on (increasing) phenomenological similarity between model and target. Two issues with this approach deserve critical discussion here.

First, phenomenological similarity is typically not treated as an empirical criterion in the literature. The level of phenomenological similarity that is considered sufficient (i.e., for which the research community "settles") currently seems to be a matter of expert consensus. For example, most may agree that a joystick movement as fear-conditioned stimulus is phenomenologically more similar to a pain patient bending his knees than a fear-conditioned geometrical figure is, but whether it comes close enough is currently treated as a matter of

opinion. Among the same vein, many would probably agree that a movie clip of a sexual assault as unconditioned stimulus is phenomenologically more similar to a sexual assault than an electric shock is (Wilhelm, 2021), but some may say that it is still not similar enough (e.g., because a movie clip is not multisensory, while an assault is) or even say that the shock is more similar (e.g., because of its tactility).

Second, and perhaps more importantly, phenomenological similarity is neither necessary nor sufficient for high translational value of a model. For example, a rat model of hallucinations may lack phenomenological similarity to the hallucinations that a schizophrenia patient reports but may still allow for successful translation to this target (e.g., if the rat reacts strongly to drugs that also reduces hallucinations in schizophrenia; Vervliet & Raes, 2013). A classic way to illustrate that phenomenological similarity is not sufficient either involves reference to horses and zebras. These animals may look alike (i.e., there is some phenomenological similarity), but what we know from riding horses does not translate to zebras (i.e., as a model for zebras, horses have little information to offer in the area of riding). This is not to say, however, that successful models should always be devoid of phenomenological similarity. Indeed, some models are phenomenologically similar to their target phenomenon – think back of the engineer who places a miniature car in a wind tunnel in order to learn about the aerodynamics of a life-sized car – and overlap in phenomenology may be an important source of inspiration when developing models (Vervliet & Raes, 2013).

As an additional argument for not putting too much emphasis on phenomenological similarity, it may be worth noting that phenomenological similarity is a criterion that cannot be applied to computational and cognitive models (i.e., it is meaningless to ask whether these models look like their target; see Box 1 and 2). Crucially, however, such models can be successful in terms of translation.

Functional similarity

In line with our argument that phenomenological similarity (e.g., an electric stimulus in a model of electrophobia) is neither necessary nor sufficient for high translational value of a model, philosophers have argued that there are forms of similarity that do not necessitate this kind of similarity. For example, there can also be a similarity in relations (e.g., the relation of a father to his children is similar to the relation of the state to the citizens). What the model then shares with its target is not (only) a set of phenomenological features, but a pattern of relations (Hesse, 1963; Stanford Encyclopedia of Philosophy, 2012). Such models can also be used to infer information about a target phenomenon (e.g., if punishing fathers create fearful children, then punishing states may create fearful citizens).

It is in this context that we propose to systematically assess the *functional similarity* between the model and target phenomenon in order to enhance translational value and to have more effective modelling practices. A test of functional similarity concerns a test of the extent to which there is overlap between variables that affect performance in the model and variables that affect the target phenomenon. Testing functional similarity can go in two directions (also see Vervliet & Raes, 2013). First, one can assess whether what is learned from the model also holds for the target phenomenon. For example, if a certain pharmacological agent reduces fear responding in the human fear conditioning paradigm, then one can test whether it also reduces real-life anxiety- and stress-related complaints. While this may seem self-evident, this empirical exercise is not always carried out (Scheveneels et al., 2016). Such tests would not only strengthen the practical impact of research – which is, after all, the holy grail of modelling – but would also serve as a more general check of whether the model has translational potential.

Second, one can assess functional similarity in the other direction and investigate whether knowledge about the target phenomenon translates to the laboratory model (for a related approach in the context of measurement optimization see Bach, Melinščak, Fleming, & Voelkle, 2020). Advantages include that the utility of the model is tested before costly

interventions in the outside world are based on it (and that it allows experimental psychologists to stay in the territory where they perform best, namely in the laboratory). For example, in a series of unpublished studies, we developed a conditioning task that aims to serve as a model of real-life grief after the loss of a loved one (i.e., the target phenomenon). In this task a neutral stimulus was paired with the name of a loved one (i.e., acquisition training), followed by presentations of the neutral stimulus by itself (i.e., extinction training). In order to assess the merits of this model, we assessed whether variables that are already known to affect the target phenomenon (e.g., more grief after losing somebody to whom one is close) affect performance in the laboratory model in a similar way (e.g., more cue-elicited craving and less extinction if the name of somebody close than if the name of somebody not close is paired with the neutral stimulus). Overlapping effects can be taken as a piece of support for the translational potential of the model (i.e., what is learned from the model is more likely to also extend to the target phenomenon) and as support that it is worthwhile to (continue to) invest in the model (but see below for a discussion of limitations). Note the difference with an approach that focuses on phenomenological similarity: In such case, one would merely try to enhance the surface similarity between model and target phenomenon (e.g., use a picture or the odor of the lost loved one instead of the name of the loved one). In contrast, functional similarity relies on the assumption that a higher number of variables that show parallel effects in the model and for the target phenomenon will increase the chance that a new and untested factor will also show parallel effects.

Note that functional similarity can also be used as a benchmark for the translational value of computational models and cognitive models (despite the obvious lack of phenomenological similarity between these models and their target phenomenon). For example, one can assess whether increasing the “salience parameter” in the computational Rescorla-Wagner model affects the output of the model in the same way as how affecting the salience of

the stimuli in a conditioning experiment (e.g., the intensity of a light or of a shock) affects behavior (see Box 1).

Functional similarity versus deep similarity

It is worthwhile to compare functional similarity with what one may term "deep similarity" or similarity in terms of underlying processes. A common assumption entails that models in psychology should rely on the (exact) same mental or behavioral process that is at work in the target phenomenon. For example, it is common to assume that both fear conditioning in the laboratory (i.e., the laboratory model) and real-life anxiety disorders (i.e., the target phenomenon) are due to association formation in memory (i.e., a mental process) or due to the pairing of neutral and aversive stimuli (i.e., a behavioral process; De Houwer, 2020). Although it is obvious that a model that relies on the same process as its target phenomenon may produce knowledge that holds for its target phenomenon, some issues deserve critical discussion here.

First, some models are not composed of mental or behavioral processes (e.g., 3D artefacts like the windmill in Box 2), restricting the range of this criterium. Second, overlap in processes is not easy to verify. Mental processes – defined as a series of information processing steps in the mind (see Box 2) – are unobservable, so verifying whether, for example, both fear conditioning in the laboratory and real-life anxiety disorders depend on the formations of associations in memory is a tremendous challenge and, accordingly, a topic of ongoing debate (Boddez, Moors, Mertens, & De Houwer, 2020; Mitchell, De Houwer, & Lovibond, 2009). Furthermore, since mental processes are unobservable, conclusions about overlap will ultimately depend on parallel effects of variables in the model and on the target phenomenon and, as such, on functional similarity. Behavioral processes – defined as the environmental events that cause behavior – are easier to control and verify in the laboratory than mental processes, but are still challenging to verify in real life (De Houwer, 2021). For example, one can verify whether fear in the laboratory is caused by stimulus pairings, but the complaints of

somebody suffering from a real life anxiety disorder may stem from various causes that we do not control, making it difficult to draw conclusions about the overlap in processes (De Houwer, 2020; Rachman, 1977).

Third, overlap in processes is not only difficult to verify, but may even be an insufficient and unnecessary condition for the utility of a model. It would be insufficient if the model and the target phenomenon share the process of interest, but differ in respects (e.g., in the stimulus material that is processed; Vervliet and Boddez, 2021) that change the outcome of this overlapping process. It would be unnecessary if a model that relies on other (behavioral or mental) processes than its target phenomenon is still a useful tool to discover new ways to affect a target phenomenon. Certain manipulations may indeed change the course of behavior irrespective of the precise nature of these processes, as different effects can have the same moderators. For example, fear caused by stimulus pairings (e.g., pairings of tone with shock) and fear caused by verbal messages (e.g., the message "dogs can bite" in case of dog phobia or that "the tone will be followed by shock") are similarly affected by a wide range of variables (for a review see Mertens, Boddez, Sevenster, Engelhard, & De Houwer, 2018). So, human fear conditioning in the laboratory can inform us about ways to affect fear that is not due to a conditioning history but due to verbal messages (so, despite the lack of overlap in terms of behavioral process). A second example of successful modeling in the absence of overlapping processes is how the computational Rescorla-Wagner model allows to make correct predictions about variables that affect target phenomena that are likely to rely on inferential reasoning processes that are beyond the scope of the Rescorla-Wagner rule (Boddez, Haesen, Baeyens, & Beckers, 2014).

What the two examples above illustrate is that the processes that make up the model and the target phenomenon do not need to be *exactly identical*, but need to be *similar enough*. Moreover, the examples illustrate that the processes are similar enough when (variables that

affect) the process underlying the model can provide information about the (variables that affect) the process underlying the target phenomenon. That is, deep similarity – so, similarity in terms of processes – is subordinate to functional similarity.

The above also highlights the importance of distinguishing between the heuristic and predictive purposes of models. The heuristic question concerns the question to what extent a model represents all the existing knowledge about a target phenomenon. This includes representing knowledge about the phenomenology of the target phenomenon and about the processes at play in the target phenomenon. The point that we made is that a model can have utility in spite of inadequate representation (Eronen & van Riel, 2015). In fact, that is the very point of modelling approaches, as models differ by definition from their target phenomena. Nonetheless, it also remains important to acknowledge that theoretical considerations about underlying processes can still be a source of inspiration when developing models, just like phenomenological similarity can be.

Limitations of functional similarity

Functional similarity has the important advantage of turning questions about the translational value of a model into empirical questions: Do variables affect the model and target phenomenon in a similar way? Despite this advantage, there are also challenges when using this approach.

First, there can be a challenge of operationalizing the variables that are used to test the functional similarity at the side of the model. The most straightforward scenario is a scenario in which the variable that affects performance in the model (e.g., fear conditioning) and the target phenomenon (e.g., anxiety disorders) is the same (e.g., a pharmacological substance). However, the variable will not always be the exact same, because an operationalization will often be used in the laboratory. For example, a researcher may assess whether the effect of a safety instruction (e.g., “from this moment onwards the chance on getting a shock is close to

zero") on subsequent extinction learning is the same as the effect of psycho-education on subsequent exposure therapy (Scheveneels, Boddez, De Ceulaer, & Hermans, 2019). In such case, the operationalization of the variable that is used to test the functional similarity – psycho-education operationalized as a safety instruction –is based on rhetorical arguments and thus not set in stone, although it may affect one's conclusions about functional similarity. The same holds for studies that suggest that specific forms of safety behavior are not as detrimental for exposure therapy (Levy & Radomsky, 2014; Rachman, Radomsky, & Shafran, 2008) as one would infer from the negative effects of operationalizations of safety behaviors in extinction studies (e.g., Craske et al., 2014).

Second and relatedly, although functional similarity comes down to an empirical question, there is no criterion to decide when a model can be ultimately accepted or rejected as a translational tool. Given that a model is never identical to its target phenomenon, maximal similarity (i.e., identity) will not be attained. However, as mentioned above, it does seem to be a reasonable working assumption that a higher number of established factors that show parallel effects in the model and in the target phenomenon will increase the chance that a new and untested factor will also show parallel effects (Pippard, 1998). There may still be exceptions though. Take the classic example of the horses and zebras that we discussed above. They may respond very differently to a rider (i.e., a variable with a different effect), but still respond similarly to another intervention (e.g., have a shiny coat after use of a certain shampoo). In such cases, the challenge becomes to delineate those areas in which functional similarity does (e.g., the area of riding) and does not apply (e.g., the area of grooming), so that one can pinpoint the areas in which the model has translational potential.

Conclusion and suggestions for research

In conclusion, we propose to assess whether variables have a similar effect in models and their target phenomena. This assessment of functional similarity may be a powerful tool to

assess the translational value of models in the field of experimental psychopathology and beyond.

With respect to future research, we recommend that researchers would first be explicit about how they delineate their model, their target phenomenon and the relation between the two. Second and most crucially, they may assess the functional similarity between models and their target phenomena before these models become a mainstay in the literature. This could be an important step to increase the translational value of research and the chances of achieving positive societal impact with (experimental) research.

References

Bach, D. R., Melinščak, F., Fleming, S. M., & Voelkle, M. C. (2020). Calibrating the experimental measurement of psychological attributes. *Nature human behaviour*, *4*(12), 1229-1235.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. London, UK: Routledge.

Beckers, T., Krypotos, A-M., Boddez, Y., Effting, M., Kindt, M. (2013). What's wrong with fear conditioning? *Biological Psychology*, *92* (1), 90-96.

Boddez, Y. (2018). The presence of your absence: A conditioning theory of grief. *Behaviour research and therapy*, *106*, 18-27.

Boddez, Y., Moors, A., Mertens, G., & De Houwer, J. (2020). Tackling fear: Beyond associative memory activation as the only determinant of fear responding. *Neuroscience & Biobehavioral Reviews*, *112*, 410-419.

Boddez, Y., Haesen, K., Baeyens, F., & Beckers, T. (2014). Selectivity in associative learning: a cognitive stage framework for blocking and cue competition phenomena. *Frontiers in psychology*, *5*, 1305.

Buabang, E. K., Boddez, Y., De Houwer, J., & Moors, A. (2021). Don't make a habit out of it: Impaired learning conditions can make goal-directed behavior seem habitual. *Motivation Science*, *7*, 252-263.

Chomsky, N. (1959). Review of B.F. Skinner, Verbal behavior. *Language*, *35*, 26-58.

De Houwer, J. (2020). Revisiting classical conditioning as a model for anxiety disorders: A conceptual analysis and brief review. *Behaviour research and therapy*, *127*, 103558.

De Houwer, J. (2021). On the challenges of cognitive psychopathology research and possible ways forward: Arguments for a pragmatic cognitive approach. *Current Opinion in Psychology*.

De Houwer, J., Barnes-Holmes, P. M. D., & Barnes-Holmes, Y. (2017). What is cognition? A functional-cognitive perspective. In S. C. Hayes & S. G. Hofmann (Eds.), *Process-based CBT* (pp. 119–136). Oakland, CA: New Harbinger Publications.

De Houwer, J., & Hughes, S. J. (2020). *The psychology of learning: an introduction from a functional-cognitive perspective*. MIT Press.

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*, 1556-1557.

Demirkaya, B., Dereli, T., & Güven, K. (2019). Analog black holes and energy extraction by super-radiance from Bose Einstein condensates (BEC) with constant density. *Heliyon*, *5*, e02497.

Derman, E. (2011). *Models. Behaving. Badly.: Why Confusing Illusion with Reality Can Lead to Disaster, on Wall Street and in Life*. Simon and Schuster.

Draaisma, D. (2000). *Metaphors of memory: A history of ideas about the mind*. Cambridge University Press.

Eronen, M., & van Riel, R. (2015). Understanding through modeling: the explanatory power of inadequate representation. *Synthese*, *192*, 3777-3780.

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

Hesse, Mary (1963), *Models and Analogies in Science*. London: Sheed and Ward.

Krypotos, A. M., & Engelhard, I. M. (2020). Pavlovian-to-instrumental transfer in subclinical obsessive–compulsive disorder. *Journal of Experimental Psychopathology*, *11*, 2043808720925244.

Kredlow, M. A., de Voogd, L., & Phelps, E. A. (2020, June 15). Laboratory Analogues and Therapy Procedures: A Case for Translation from the Clinic to the Laboratory. https://doi.org/10.31234/osf.io/7r6p2

Kuorikoski, J., & Ylikoski, P. (2015). External representations and scientific understanding. *Synthese*, *192*, 3817-3837.

Levy, H. C., & Radomsky, A. S. (2014). Safety behaviour enhances the acceptability of exposure. *Cognitive Behaviour Therapy, 43*, 83-92.

Lewis Jr, N. A., & Wai, J. (2021). Communicating what we know and what isn't so: Science communication in psychology. *Perspectives on Psychological Science*.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Meulders, A. (2020). Fear in the context of pain: Lessons learned from 100 years of fear conditioning research. *Behaviour research and therapy*, *131*, 103635.

Mitchell, C. J., De Houwer, J., & Lovibond, P.F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*, 183-198.

Moors, A. (2007). Can cognitive methods be used to study the unique aspect of emotion: An appraisal theorist's answer. *Cognition and Emotion, 21,* 1238-1269

Mertens, G., Boddez, Y., Sevenster, D., Engelhard, I. M., & De Houwer, J. (2018). A review on the effects of verbal instructions in human fear conditioning: Empirical findings, theoretical considerations, and future directions. *Biological Psychology*, *137*, 49-64.

Pal, S., Kabir, S. M. H., & Talukder, M. M. M. (2015, November). Aerodynamic analysis of a concept car model. In *Proceedings of the International Conference on Mechanical Engineering and Renewable Energy*.

Pittig, A., van den Berg, L., & Vervliet, B. (2016). The key role of extinction learning in anxiety disorders: behavioral strategies to enhance exposure-based treatments. *Current opinion in psychiatry*, *29*, 39-47.

Rachman, S. J. (1977). The conditioning theory of fear acquisition: a critical examination. *Behaviour Research and Therapy, 15*, 375–387.

Rachman, S., Radomsky, A. S., & Shafran, R. (2008). Safety behaviour: A reconsideration. *Behaviour research and therapy, 46*, 163-173.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Scheveneels, S., Boddez, Y., De Ceulaer, T., & Hermans, D. (2019). Ruining the surprise: The effect of safety information before extinction on return of fear. *Journal of Behavior Therapy and Experimental Psychiatry*, *63*, 73-78.

Scheveneels, S., Boddez, Y., & Hermans, D. (2021). Predicting clinical outcomes via human fear conditioning: A narrative review. *Behaviour Research and Therapy*, *142*, 103870.

Scheveneels, S., Boddez, Y., Vervliet, B., & Hermans, D. (2016). The validity of laboratory-based treatment research: Bridging the gap between fear extinction and exposure treatment. *Behaviour Research and Therapy*, *86*, 87-94.

Stanford Encyclopedia of Philosophy (2012). Models in science. https://stanford.library.sydney.edu.au/archives/fall2016/entries/models-science/

Stanford Encyclopedia of Philosophy (2019). Natural Properties. https://plato.stanford.edu/entries/natural-properties/

Törneke, N. (2010). *Learning RFT: An introduction to relational frame theory and its clinical application*. New Harbinger Publications.

Vervliet, B., & Boddez, Y. (2020). Memories of 100 years of human fear conditioning research and expectations for its future. *Behaviour Research and Therapy*, 103732.

Vervliet, B., & Raes, F. (2013). Criteria of validity in experimental psychopathology: application to models of anxiety and depression. *Psychological medicine*, *43*, 2241-2244.

Waters, A., LeBeau, R., & Craske, M. (2017). Experimental psychopathology and clinical psychology: An integrative model. *Psychopathology Review, 4*, 112-128. doi:10.5127/pr.038015

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology, 3*, 1–14.

Weisberg, M. (2013). *Simulation and Similarity: Using models to understand the world.* Oxford: Oxford University Press.

Wilhelm, F. (2021). *Trauma-film conditioning as an experimental psychopathology model for posttraumatic stress disorder and anxiety disorders: what have we learned in 10 years of research?* Presented at the European Meeting on Human Fear Conditioning, Online, 17 - 19 May 2021.

Zenses, A. K., Lenaert, B., Peigneux, P., Beckers, T., & Boddez, Y. (2020). Sleep deprivation increases threat beliefs in human fear conditioning. *Journal of sleep research*, *29*, e12873.

Footnotes

[1]Other authors (Derman, 2011) have tried to clarify the meaning of the term model by contrasting it with the term theory. Theories make a (right or wrong) statement about what something is. Models differ by definition from their target phenomenon (and are therefore always "wrong" or, in other words, their heuristic representation of the target phenomenon is always inadequate; Eronen & Van Riel; 2015). Still, models may be used to tell us something about a target phenomenon in the way a metaphor can be used to tell something about a phenomenon to which it is not literally applicable.

[2]Phenomenological similarity may remind readers of face validity, while functional similarity and deep similarity may remind readers of predictive validity and construct validity, respectively (e.g., Vervliet & Raes, 2013). Still, we believe that there are good reasons for relying on a terminology that invokes similarity relations. First, a model may bear a specific similarity relation to one target phenomenon, but not to another target phenomenon. In our framework, when evaluating a model, it is therefore all about the relations between the model and the target phenomenon at hand. Second, saying that a model "has [face, predictive or construct] validity" may seem to imply that the decision on the matter is final and conclusive. This may lead to unfounded trust in the model. In our framework, similarity relations remain under investigation. More precisely, we especially argue for the continuous assessment of functional similarity relations, as it is a plausible working assumption that more factors that show a parallel effect in the model and on the target phenomenon will increase the chances that a new and untested factor will also translate from model to target phenomenon. On top of this, we refer to the reader to the main text for a discussion of problems to reach consensus and final conclusions about face and construct validity.