American Behavioral Scientist
2023, Vol. 67(2) 311–331

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00027642211066055
journals.sagepub.com/home/abs

Disruption of Social Orders in Societal Transitions as Affective Control of Uncertainty

Jesse Hoey[1] and Tobias Schröder[2]

Abstract

Bayesian affect control theory is a model of affect-driven social interaction under conditions of uncertainty. In this paper, we investigate how the operationalization of uncertainty in the model can be related to the disruption of social orders—societal pressures to adapt to ongoing environmental and technological change. First, we study the theoretical tradeoffs between three kinds of uncertainty as groups navigate external problems: validity (the predictability of the environment, including of other agents), coherence (the predictability of interpersonal affective dynamics), and dependence (the predictability of affective meanings). Second, we discuss how these uncertainty tradeoffs are related to contemporary political conflict and polarization in the context of societal transitions. To illustrate the potential of our model to analyze the socio-emotional consequences of uncertainty, we present a simulation of diverging individual affective meanings of occupational identities under uncertainty in a climate change mitigation scenario based on events in Germany. Finally, we sketch a possible research agenda to substantiate the novel, but yet mostly conjectural, ideas put forward in this paper.

Keywords

Bayesian affect control theory, social order, societal change, uncertainty

[1]University of Waterloo, Waterloo, ON, Canada
[2]Potsdam University of Applied Sciences, Potsdam, Germany

Corresponding Author:
Jesse Hoey, Cheriton School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada.
Email: jhoey@uwaterloo.ca

Introduction

Affect Control Theory (ACT) is a theory of how the social order is enacted and maintained in everyday interactions through affective processes modeled with sentiments. The theory is based on the assumption of static cultural sentiments about identities and occupational status (Heise, 2007; Freeland & Hoey, 2018), the temporal dynamics of these sentiments, and the control principle that optimizes over the difference between in-context sentiments (impressions) with out-of-context sentiments. That is, most ACT research assumes that culture, including its dynamics, is fixed at a point in time. BayesACT, a probabilistic generalization of ACT, removes this restriction, and allows sentiments, dynamics and control to be learnable. Thus, BayesACT allows for, and can adapt to, changes in social relations (Hoey et al., 2016; Schröder et al., 2016; MacKinnon & Hoey, 2021; Hoey et al., 2021). The result is a mathematical tool that can be used to simulate shifts in social orders, and thereby explain emotional reactions to major changes, including emergent political conflict and polarization of belief systems.

Such a tool is much needed given the profound transformations many societies are currently experiencing. While contemporary human civilization enjoys unprecedented wealth and progress (Pinker, 2018), it is struggling to adapt to rapid changes of the ecological and technological environment. Transgressions of the planetary boundaries of resources threaten the biological living conditions for many species on earth, possibly including our own (cf. Rockström et al., 2009). Evolving digital technology and infrastructure may help to deal with ecological problems (e.g., through better management of energy infrastructures) but at the same time cause their own adaptive challenges, disrupting entire industries. As MacKinnon and Heise (2010) point out from an ACT perspective, economic and occupational institutions are central for the structure of selves and the social order in contemporary societies. Therefore, when current societal transformations increase uncertainty about the meaning of occupational identities, this will have profound consequences for the affective control of social interactions (think of coal miners or petroleum engineers becoming obsolete in a renewable-energy world, or specialized doctors such as radiologists becoming replaced by machine learning algorithms). It is the goal of the present paper to make a theoretical proposal for how the BayesACT model can be used as an analytical tool to understand these changes.

We start by briefly introducing BayesACT, leaving out details that are extensively covered in previous papers (Hoey et al., 2016, 2021; MacKinnon & Hoey, 2021; Schröder et al., 2016) and instead focusing on three different kinds of uncertainty in the BayesACT model and the tradeoffs between them. Next, we discuss how the management of uncertainty as operationalized in BayesACT relates to social conflict accompanying rapid ecological and technological change. As an example, we then turn to a simulated interaction of a coal miner with an environmentalist to illustrate the promise of our approach to better understanding the social/political psychology of societal transformation from an affect control theory perspective. We finish with a brief discussion of a possible research agenda to empirically test these novel theoretical ideas and hint at some practical implications for managing change in organizations and society. While management of uncertainty is understood to play a major role in human interaction (FeldmanHall & Shenhav, 2019; Hirsh et al., 2012), precise mathematical models of this role are scarce. Here we propose a possible roadmap for this endeavor, which we believe may be a useful bridge between social psychological reasoning and Bayesian inference methods from artificial intelligence.

Bayesian Affect Control Theory

Affect Control Theory (ACT; Heise, 2007) arises from symbolic interactionism, according to which people base their decisions and actions to some degree on culturally shared meanings of things, which have evolved and are maintained and constantly reproduced in social interaction. When socialized in a given culture, people are assumed to have internalized semantic structures that serve as default frames to keep their individual decisions in line with social expectations. ACT proposes a fundamental link between denotative representations of social objects (e.g., linguistic labels for identities, such as "doctor" or "nurse") and the associated connotative meanings, which are measured using semantic differential scales in an affective space spanned by dimensions of evaluation, potency, and activity (EPA; Osgood, 1969). Dictionaries of identity concepts along with their empirical measures of EPA connotations from population surveys have been interpreted as collective representations of the social order, with an analytical focus of empirical research on the commonality, consensus, and stability among members of a linguistic community (e.g., Ambrasat et al., 2014).

Affect Control Theory includes a control mechanism, which is based on the difference between fundamental, out-of-context sentiments and the in-context transient impressions that are formed when social events occur. The *affect control principle* is to find the optimal alignment between these two sets of sentiments. This optimization reveals the connotative meanings of the most emotionally aligned action for an agent to take next. The dynamics of impression formation are also measured, by asking participants about specific *actor-behavior-object* (ABO) situations, e.g., *librarian* (EPA: {1.8,-0.3,-2.1}) *reprimand* (EPA:{-0.4,1.6,1.1}) *bookworm* (EPA:{1.6,0.3,-2.3), and asking them to rate each element.[1] The basic premise of ACT is that such situations are assessed in a unified way. The difference between this combined estimate and the out-of-context estimates (fundamental sentiments), used as an optimization loss, guides agents' decisions about actions or reinterpretations of the situation to reduce emotional incoherence.

Bayesian Affect Control Theory Fundamentals

Affect Control Theory can be used to simulate an interaction *given* a set of identity and behavior labels. The interaction modeled by ACT is referred to as connotative, as it gives the affective meanings of events. In BayesACT, the restriction of fixed identity and behavior labels is lifted, and sentiment variance is also taken into account (Hoey et al., 2016, 2021; Schröder et al., 2016). Consequently, BayesACT must model the identities and behaviors denotatively as well as connotatively, as it must maintain a distribution over them representing the agent's belief that these are the current identities and behaviors at play. The denotative component models the linguistic labels representing the state of the world (e.g., the identity of "doctor" or "nurse," or the behavior of "advising" or "ignoring"), while the connotative component is like ACT but includes sentiment distributions. BayesACT can be thought of as maintaining a *frame* that defines the situation in terms of a set of propositions and their associated sentiments. These *frames* are also referred to as *state spaces*: an enumeration of all possible values that can be taken on. For example, the denotative state space over identities consists of all known labels for identities (e.g., "nurse," "doctor," and "administrator"). The connotative state space is the real-valued space of evaluation, potency and activity as in ACT.

The denotative component can encode constraints on the denotative state. Such constraints constitute a toolkit of social norms, rules, expectations of behavior, or other aspects of the context, and may include denotative stereotypes (e.g., associations between characteristics of people and expected behaviors). Actions may also be constrained in this way, leading to hard-coded policies of action sometimes referred to as *repertoires*, patterns of behaviors that are pointed to and labeled as *narratives,* and sets of repertoires and narratives that form *institutions* (MacKinnon & Heise, 2010; Vaisey & Valentino, 2018). For example, it may be connotatively coherent for a "coal miner" to "work in a coal mine," but this may not be possible because the mine was shut down for being too polluting. Such constraints may be uncertain, and thus may be modeled with a probability function. Overall, BayesACT generalizes ACT by explicitly representing the distribution over sentiments in a two-level partially observable Markov decision process (Åström, 1965).

In the BayesACT model, there are thus two levels of representation and control. One, connotative level represents the sentiments associated with a given identity-behavior event. A connotative action is explicitly encoded as the behavior sentiment. The other, denotative level represents the actual event itself, including the identity and behavior labels. Denotative actions are encoded as behavior labels. The *somatic transform* is a probabilistic function that measures the congruence between denotative and connotative interpretation as a probability distribution over labels and a density function in EPA space, respectively (MacKinnon & Hoey, 2021). The somatic transform allows BayesACT to operate as a *dual-process* model (Hoey, 2021; Hoey et al., 2021; Vaisey, 2009) in which actions are optimized for both connotative coherence and denotative objectives (e.g., goals). It has an equivalence in non-Bayesian ACT as a *sentiment dictionary,* which defines *symbolic boundaries* (Vaisey & Valentino, 2018), but is fixed and non-probabilistic.

At any time, BayesACT maintains a probability distribution over both the connotative space and the denotative space. One can imagine this distribution in the connotative space as a continuous function in three dimensions, higher values of which represent more likely connotative meanings for the current state. In the denotative space, the distribution is a multinomial over possible labels (e.g., a person in a hospital setting who is unknown to you may be identified as a "doctor" or an "intern" with nearly equal probability, as a "nurse" with a smaller probability, or as a "patient" with a very small probability, which would be represented as a set of four numbers that sum to 1, e.g., [0.45,0.34,0.01,0.2]). These probability distributions can be characterized by how *dispersed* they are. A very dispersed distribution is one that is spread evenly over the state space. The opposite of this is a precise distribution, in which the probability mass is concentrated on only a few denotative options, or in a small region of the connotative space. One can think of these as the variance in the prior and posterior distributions over connotative and denotative state spaces (frames). Figure 1a shows example distributions in sentiment space for one dimension for these four example identities, as well as the denotative distribution as a bar chart in Figure 1b and the resulting connotative mixture distribution in Figure 1c.

In the BayesACT model, one denotative action is randomly selected from the denotative probability distribution over behaviors for enactment in the following event. Thus, the action selected is a product of denotative and connotative reasoning and there are two control mechanisms used to select actions that are intertwined. As the connotative and denotative spaces have different dynamics (predictor functions), it rests on the relative strengths of the predictions in each, coupled with the relative strength of the somatic transform to determine which of these systems will carry more "weight" in the determination of action. Thus, the two control mechanisms are tightly connected (*inextricable*), and *complementary* in that one takes over when the other fails, and the control policy associated with each optimizes different objective functions (MacKinnon & Hoey, 2021). While the denotative state optimizes over preferences, the connotative policy optimizes shared emotional meanings. The principle of *complementarity* states that both are necessary for action to take place. While denotative framings guide decision-theoretically rational decision-making processes, connotative meanings *frame* the denotative process, giving rise to deep emotional constraints on the possible actions that are considered.

Management of Uncertainty in Bayesian Affect Control Theory

There are three primary elements in the BayesACT model that represent uncertainty: the dispersion of the denotative distribution, the dispersion of the connotative distribution, and the dispersion of the connection between the two. First, consider the denotative model. The dispersion in an actor's model is directly related to its resource bound (its ability to carry out computations) and to the complexity of the environment. Environments with many other actors attempting to cooperate are highly complex, and therefore contain much more ambiguity arising from the actor's inability to model the large quantity of potential configurations that its known environment could be in. An actor can either accept the increased ambiguity, and accept dealing with an *invalid* world (Kahneman & Klein, 2009),[2] or can allocate cognitive resources to the situation.

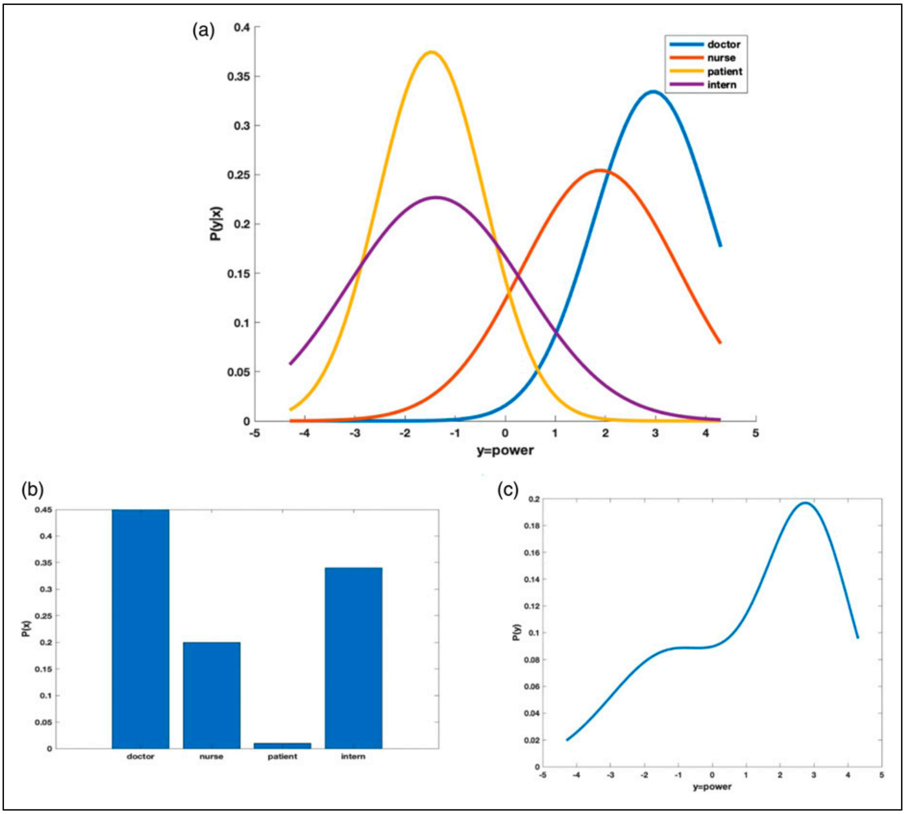

Figure 1. Example of connotative and denotative distributions. (a) Normal distributions on the Power/Potency dimension estimated from the USA 2015 survey (Smith-Lovin, et al., 2016), for four different hospital-related identities. Doctors are typically viewed as having more power than nurses, patients and interns. Sentiments about nurses and interns have larger variance than those about patients and doctors, for which there is more consensus. (b) Example denotative distribution over these four identities. (c) Resulting mixture distribution showing the sentiments attributed to this person given the denotative distribution shown in (b).

This latter option corresponds to gaining expertise in the domain, and so the actor would be labeled an "expert," and the situation would become more *valid* from their perspective (but maybe not from others in the same environment). Thus, low-complexity or high-resource environments are considered valid, while high-complexity or low-resource environments are considered invalid (Kahneman & Klein, 2009). The validity of the denotative model is represented by a parameter. $\hat{\delta}$[3] In a simple two-state situation (e.g., two labels in a vector ["nurse," and "doctor"]), the denotative distribution might be invalid (e.g., a distribution over the same vector space $[\widehat{\boldsymbol{\delta}}, 1 - \widehat{\boldsymbol{\delta}}] = [0.5, 0.5]$, so there is great uncertainty about which label should be applied to this person), or valid (e.g., $[\widehat{\boldsymbol{\delta}}, 1 - \widehat{\boldsymbol{\delta}}] = [0.99, 0.01]$, so the person is almost certainly a "nurse," not a "doctor").

The connotative model's precision is afforded by the strength of the affect control principle, which states that persons will behave in ways so as to increase *coherence* between their feelings about entities out-of-context (fundamental sentiments) and in-context (transient impressions). The strength of the affect control principle is represented by a parameter ($\widehat{\boldsymbol{\alpha}}$), which was arbitrary in ACT (Heise, 2007), but obtains scale through comparison with validity (denotative precision) and the strength of the somatic transform ($\widehat{\boldsymbol{\gamma}}$), (Hoey et al., 2021). The parameter $\widehat{\boldsymbol{\gamma}}$ measures *degree of inextricability* of denotative and connotative meaning spaces, and makes up the third element in the representation of uncertainty, which we will here refer to as *dependence*; that is, the degree to which connotative meanings depend on denotative meanings, and vice versa. Agents in more diverse environments will have weaker somatic transforms (smaller $\widehat{\boldsymbol{\gamma}}$), and weaker affect control principles (smaller $\widehat{\boldsymbol{\alpha}}$), as they are confronted with a greater diversity of different identities and behaviors.

Imagine you are in a hospital and you see someone wearing a white coat. The validity is how certain you are that this is, in fact, a "doctor" (not an "intern"). If it is your family doctor, this certainty may be very high (uncertainty in the label is very low). If it is someone you don't know, your certainty may be lower. Given that this person is a doctor, your uncertainty in the sentiments about the doctor in EPA space is given by the dependence parameter. The lower this parameter, the less sure you are about how your culture feels about doctors. Perhaps you live in a place where there are "good" doctors and "bad" doctors in roughly equal proportions. Your certainty about how to interpret this doctor will be much lower than if you live in a place where almost all doctors are considered "good" by society at large. Finally, the coherence of the situation is how certain you are that this doctor, who is good (say), will behave in a way that is consistent with his and your identities.

To sum up, the three elements are

- *validity*: the precision of the denotative model (determines the denotative distribution, includes denotative evidence, parameter $\widehat{\boldsymbol{\delta}}$);
- *coherence*: the precision of the affect control principle (determines the connotative distribution, includes connotative evidence and dynamics[4], parameter $\widehat{\boldsymbol{\alpha}}$);
- *dependence*: the precision of the somatic transform, or the degree of inextricability (connects connotative and denotative, parameter $\widehat{\boldsymbol{\gamma}}$).

Now, consider how these three elements will trade off against each other. The distribution over both connotative and denotative state spaces is the combination of these three representations of uncertainty. One can think of these precisions as the strength of constraints in both connotative and denotative levels, and between them. In BayesACT, a strong set of constraints at one level will mean that the elements of the state (denotative or connotative) connected to those constraints will have a larger influence and be more well defined. Thus, a precise affect control principle means that transient impressions and fundamental sentiments are strongly constrained to be close to one another. If the other distributions (to the denotative state and associated denotative evidence) are kept the same, then an increase in affect control precision will strengthen associations of sentiment to their culturally accepted prescriptions as described by ACT, probabilistically re-evaluating denotative evidence if necessary. On the other hand, a precise denotative distribution will be skewed in one direction or the other, and will more heavily shift the connotative meanings to be in line with one denotative interpretation of the state.

For example, in our hospital setting, suppose you see a woman in white lab coat. There are denotative constraints (with strength $\widehat{\boldsymbol{\delta}}$) related to who is likely to be in a hospital (nurses, doctors, patients and interns), and who is likely to be wearing a white lab coat (doctors and scientists). These constraints may be satisfied (the person in the lab coat giving orders *is* a doctor), or not (the person is actually an intern). There are also connotative constraints (with strength $\widehat{\boldsymbol{\alpha}}$) that indicate what types of actions are expected by these identities (people will be deferential to doctors because they are good, and doctors will direct others because they are powerful, whereas other workers may take direction and do not require as much deference). Further, stereotypes based on physical traits (e.g., gender) may skew expectations denotatively (the posterior probability that the woman in the white coat is a nurse may be higher due to a gender stereotype), or connotatively (towards less deference culturally associated with the fixed and observable gender characteristics). Finally, denotative and connotative factors are constrained to be consistent through the somatic transform (with strength given by $\widehat{\boldsymbol{\gamma}}$), (e.g., doctors are expected to be more powerful than nurses). While the woman in the white coat may be estimated to be a nurse denotatively, she may be estimated to be a doctor connotatively because she is ordering people around. These two interpretations would conflict to a degree given by $\widehat{\boldsymbol{\gamma}}$. All three of these elements combine to form a final determination, in which the denotative identity of the person in the lab coat may be more weighted towards doctor than nurse. In situations of conflict, the determination can still be made, although the final distribution may be much more dispersed due to the disagreement.

The three aspects of uncertainty we have been discussing are modeled with parameters in BayesACT ($\widehat{\boldsymbol{\delta}}$,$\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\gamma}}$) that control the amount of dispersion (variance) in distributions of connotative and denotative states. Take a moment to consider what these dispersions really mean. Using the "artificial intelligence" method of seeing if we can build something that works like a human, these variances are the degree of uncertainty in the agent's "mind." However, that degree of uncertainty is directly predicated on the degree of uncertainty in this agent's ecological niche, *and* vice versa. In this sense, we can equate individual level variance with cultural level variance, but only if there is sufficient coordination amongst the agents that share the culture. Without this coordination (e.g., everyone working on a joint project), the mapping becomes much weaker.

The analysis of the three uncertainty parameters proceeds by first establishing that they cannot all three be small (very dispersed). Should this happen, both the connotative and denotative distributions of each agent in the group, and therefore in the group as a whole, are guaranteed to also be very dispersed. Why is this a problem? Recall that both connotative and denotative elements in BayesACT contain a representation of action, and can be used to derive a policy of action that an agent is consciously aware of. For an agent to be *motivated* to take an action according to this policy, the distribution over actions must be sufficiently precise. The precision of the distribution over connotative actions then provides the impetus to act. This fits well with neuroscientific evidence that the strength of emotional appraisals (what we call the connotative state) are instrumental in catalyzing action (Damasio, 1994; Gilead et al., 2021). Therefore, if all agents in a group have very dispersed connotative distributions, all will be less certain of risk, and none will be motivated to act, leading to a dysfunctional group.[5] Can all these dispersions be very small (high precision)? The answer is unfortunately "no," because the posterior distributions have to be located in somewhat the same region of the state space in order to "find" each other. Using the notion of the three sets of constraints above, the system is over-constrained, which often leads to the non-existence (or great difficulty of finding) a solution. Therefore, we see that only one or two of these constraints can be strong at one time.

Finally, we can postulate that only one of these constraints can be strong at a time. Suppose the denotative distribution was very precise, then fundamental sentiments and transient impressions would be required to be such that they agreed with this denotative interpretation, but could not be so if they were too constrained to either lie near each other. An allegorical presentation of this matching problem is that of a triangular enclosure with a mass at the center connected by a spring of a different stiffness to each corner, as shown in Figure 2.

The spring connecting the mass to each corner has a *stiffness* which is proportional to the precision of the corresponding uncertainty management element (coherence, validity, or dependence). That is, a stiffer spring translates to a more precise (= less dispersed) element. When one of these forces is relaxed (e.g., denotative uncertainty is increased, validity is decreased), the other "takes up the slack" and naturally contracts (e.g., the affect control principle is strengthened, and connotative coherence is increased in importance). To summarize the argument at this point, an increase in one form of uncertainty will force a shift in the balance of how the other forms of uncertainty are managed. In the next section, we will argue that this mechanism can explain emotionalization and polarization of social conflict in times of societal transitions.

Disruption of Social Orders

We will now turn to a discussion of the consequences of the described tradeoffs in uncertainty management as postulated by the BayesACT model in situations where the established social order is undergoing a transformation as a result of adaptive pressures

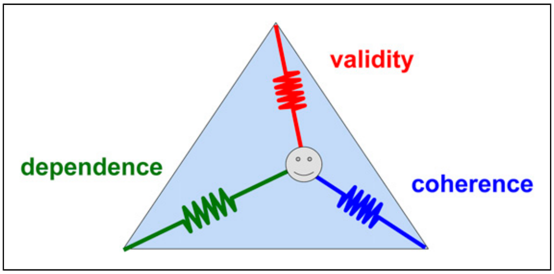

Figure 2. Spring-mass allegory for the management of uncertainty in three mutually exclusive directions. The coherence factor in this case is less dispersed (stronger) than the other two.

from ecological challenges and technological advances. While there are many contemporary contestations of the social order, we will focus here on the emergence of a digitalized sustainable society (or, more colloquially: *smart green society*), envisioned as a 21st century variant of capitalism that leverages information and communication technology to operate within the planetary boundaries of ecological resources (e.g., Rockström et al., 2009; WBGU, 2019).

The smart green transition of society inflicts considerable uncertainty on many people's social interactions, as it causes a profound disruption of the symbol system which people use to make sense of social structures. In some sense, discourse external to existing groups forces upon them a change in the social structure. This is most obvious in the case of the changing meanings of occupational identities, which are, in Western societies, one of the most important sources of social meaning and esteem (Freeland & Hoey, 2018; MacKinnon & Heise, 2010). The symbolic system of identities, also called a "theory of people" (MacKinnon & Heise, 2010), constitutes a basic belief system, which is known from neuroscientific research as a protective factor against the anxieties resulting from a state of "psychological entropy" (Hirsh et al., 2012). When entropy is high, uncertainty interferes with individuals' ability to make decisions, resulting in physiological stress symptoms. In terms of the BayesACT uncertainty parameters, changes in the "theory of people" can decrease the denotative validity (e.g., by inventing novel occupations such as *interface designer*, or *data steward*), but can also decrease the dependence between denotative and connotative meanings, and can decrease the connotative coherence.

In the case of dependence, society might force upon a group an undesired shift in affective meaning. We believe this is the case in, and a major cause of much resistance towards, the discussion surrounding climate change. Consider the example of coal or lignite miners, an occupational identity emblematic for the wealth related to industrialization and a source of pride for generations of workers. Much attention has been on the economic disruption entailed in the transition to more environmentally sustainable forms of energy, but there is also an important -and possibly underestimated-effect on affective meaning. Discussions of coal and lignite as "dirty" sources of energy become negatively attached metaphorically to the meaning of the miner identity. Portrayals of to-be-laid-off miners as in need of government help reduce their perceived potency. It is understandable if workers resist redefinitions of their once proud (i.e., positive and potent) identities at the connotative level as negative and weak.[6] Arguably, the resulting decrease of somatic-transform precision (dependence) is powerful fuel to the heated debate around economic austerity in support of climate change mitigation and explains the slow (from an environmental perspective) progress. Fukuyama (2018) similarly notes that "[a] great deal of what we conventionally take to be economic motivation driven by material needs or desires is in fact a thymotic *(roughly 'recognition by others')* desire for recognition of one's dignity or status." (italics added, p. 81).

The digital transformation of the economy causes similar increases of the uncertainty of affective meaning. This is obvious with regard to the wide-spread fears that technology will render many jobs and, hence, sources of both income and social status obsolete. But there are also more subtle changes in the social order brought forward as technology transforms the workplace. Digitalized and platform-based forms of work may enable (or cause) more egalitarian collaboration patterns with shifts of status and power from formal leadership roles to situation-specific and expertise-based identities (e.g., Zöller et al., 2020). While more egalitarian workplace interactions are widely popular, recent work based on affect control theory showed that they also tend to increase uncertainty and affective deflection due to the constant negotiation of role relations that occurs in the absence of clear hierarchies in task groups (Hoey et al., 2018; Morgan et al., 2021).

The global political challenges arising from ecological pressures and technological progress are more difficult to address in a climate of wide-spread political polarization. We conjecture that at least part of this polarization directly results from the uncertainty tradeoffs in situations of decreased validity of the denotative environment and consequences for the connotative social structure. This idea is in line with experiments showing how the effects of increased ambiguity may result in more political polarization (Bail et al., 2018). That is, if I believe in A, and you believe in B, then so long as I am not aware of you, my posterior denotative distribution may be very precise over A only. The introduction of B now expands the space of possibilities, and naturally expands the distribution to also have some mass over B (given your belief in B that I now have learned about). This increase in dispersion in the denotative state is compensated for by an increase in precision in affective meanings (i.e., dependence), leading to stronger beliefs in existing biases such as identity and stereotypes.

Illustrative Simulation: The Obsolete Coal Miner

In this section, we describe a few simulations with the BayesACT model of a scenario intended to illustrate the interplay of contested identity meanings and agents' different styles of uncertainty management in the context of a societal transition. The simulations are very simple, but we hope they will help to illustrate the promise of an affect control theoretical analysis of societal challenges. Most importantly, they are the first demonstration of the impact of the three kinds of uncertainty operationalized in BayesACT (validity, dependence, coherence) on social interactions. In contrast, previous simulations of social behaviors with affect control theory have focused entirely on the meanings of identities, behaviors, and settings as well as on impression-formation dynamics, neglecting tradeoffs in management of uncertainty. Then, we will sketch possible next steps for a systematic research agenda aimed at rigorously testing the novel, exploratory ideas in this paper.

Rationale of the Simulations

The scenario is an interaction between an environmentalist and a coal miner. Substituting coal-based electricity generation with renewable and emission-free solar energy is a major requirement in order to stay within the bounds of the carbon budget of 1.5 degree Celsius rise of average global surface temperatures as per the Paris agreement on climate change. However, the debate about phasing out coal as a source of energy is often far from factual and rational, and is subject to much political polarization. The goal of our simulations is to show how such polarization can arise from identity-verification mechanisms as theorized in Bayesian affect control theory, but that the extent to which conflict occurs in such interactions will depend on the configuration of uncertainty-management processes of the involved agents.

Formally, we are engaging in a computational experiment, where the scenarios differ in the configuration of two of the agents' uncertainty parameters (incoherence of interactions α, and independence on affective meanings γ), while the denotative invalidity δ, and the identity EPA meanings and impression-change dynamics (parameters typically varied in affect-control theory research) are the same across the simulations.[10] We study the emerging behavior dynamics in the BayesACT simulation and interpret them in terms of our narrative of societal change due to climate change mitigation policies.

Data Sources and Setup

There are four simulations in total. In each case there are two agents called Hank and Tom, who start out with two identities with equal probability of p = .5. One of the identities is that of a *citizen* for both agents, to implement in the simulation the French Republican ideal that every person regardless of what their other identities may be is a *citoyen* capable and willing to engage in democratic and rational debates. However, Hank and Tom differ in their other identity, to implement in the simulation the possibility of contestation and conflict. Hank sees himself as an *environmentalist* and Tom as a *culprit*, whom he holds to some extent accountable for the negative effects of climate change. Conversely, Tom sees himself as a *coal miner* and Hank as a *prosecutor*, who illegitimately makes moral judgments about his (Tom's) lifestyle.

For the following simulations, the impression change dynamics are based on the USA 1978 study (Schneider, 2006), while the identity and behavior sentiments (means only as variances are set using γ—see below) are taken from the USA 2015 study (Smith-Lovin et al., 2016). Both are samples of Americans at large universities. To avoid institutionally awkward simulation outcomes, which often occur in affect control theory models without institutional/denotative filters, we manually restricted the list of identities and behaviors to a set we intuited might reasonably occur in our simulated scenario. Mostly, this meant picking relevant occupational identities and behaviors, but also a few which might make sense in a more metaphorical way. For example, Tom viewing Hank as a *prosecutor* is not institutionally viable in a strict sense, but we think this choice of identity expresses the feeling of being morally "under siege" experienced by some people whose lifestyles are criticized in environmental discourse. We selected a subset of 116 identities and 29 behaviors from the main study sample by manually filtering the dictionary and only keeping words that could be logically applied to the situation at hand.[7]

In the following we show two simulations. In both, we set $\delta = 0.1$, meaning that each behavior passed from one agent to other is corrupted by Gaussian noise with a standard deviation (SD) of 0.1 in EPA space. This may or may not cause a change in the denotative label (likely not). In the first, we show agents who have *aligned* parameters: they match in size (both agents use $\alpha = 0.1$, $\gamma = 0.15$).[8] In the second simulation, we use parameters that are *mis-aligned*: one is larger (around 0.1) while the other is smaller (around 0.01), and which is which different between the two agents (so one is $\alpha = 0.01$ $\gamma = 0.25$ while the other is $\alpha = 0.5$ $\gamma = 0.01$).

Results

Table 1 shows the result of a simulation with *aligned* agents. The "d" columns show the deflection, while the "Δf" column shows the difference in the two agents' interpretation of the situation as far as sentiments go.[9] These agents are able to fairly reliably "find" the deflection-minimizing global solution where they are both *citizens* and know the other is a *citizen*. Behaviors that result in such a case are things such as "*reward*."

Even in the case of a perturbation as shown in Table 1, the agents are able to recover.

Table 2 shows the result of a simulation with *mis-aligned* agents. In this case, the agents are less reliably able to "find" each other (they still do sometimes, just less often). Behaviors that result in such a case are more accusatory, such as *discipline,* repentant, such as *obey,* or reparative, such as *pacify.* These behaviors are more relational as they directly relate to emotional factors in the interaction (who is leading whom), while the behaviors in the aligned case tend to be more neutral and may lead to more constructive interactions. We note clearly that these results are "cherry-picked" in the sense that we are only starting to scratch the surface of the rich space of models spanned by BayesACT. We are noting here only that this parameter space goes beyond variation in individual sentiments studied in ACT, and moves towards a view that integrates social forces through shared uncertainty management mechanisms that can have a significant impact on the way in which difficult relationships are managed.

Table 1. ALIGNED AGENTS. The distributions over denotative labels are cut off when the cumulative probability exceeds 0.6. d= deflection, f fundamental difference (smaller is better), Here the agents have parameters (Hank) $\alpha = 0.1$, $\beta = 0.01$, $\gamma = 0.15$ and (Tom) $\alpha = 0.1$, $\beta = 0.01$, $\gamma = 0.15$. Also, here we "force" the first action to be "test," aligned with the prosecutor-culprit pair and you can see the effect of the perturbation—it persists for about three iterations, and then is lost after multiple rounds of "appeasement."

Hank											Tom	
	Hank		Tom		Behav	Tom		Hank				
d	id	p	id	p	↔	Id	p	id	p	d	Δf
0.0	0.5	Citizen	0.5	Citizen	Test	0.5	Citizen	0.5	Citizen	0.0	4.0
	0.5	Environmentalist	0.5	Culprit	→	0.5	Coal miner	0.5	Prosecutor		
3.0	0.7	Environmentalist	0.9	Culprit	Appease	0.4	Coal miner	0.8	Prosecutor	1.1	2.5
					←	0.1	Ciner	0.0	Colleague		
						0.1	Caborer				
1.9	0.6	Colleague	0.6	Crane operator	Pacify	0.3	Coal miner	0.9	Prosecutor	3.0	3.0
					→	0.3	Ciner				
6.0	0.5	Environmentalist	0.6	Culprit	Reward	0.5	Corker	0.5	Crane operator	1.9	0.2
	0.2	Peer			←	0.2	Crane operator	0.2	Citizen		
2.1	0.4	Citizen	0.4	Citizen	Back	0.3	Citizen	0.3	Citizen	1.9	0.2
	0.2	Colleague	0.3	Crane operator	→	0.2	Colleague	0.2	Colleague		
						0.1	Corker	0.2	Crane operator		
2.0	0.4	Citizen	0.4	Citizen	Reward	0.4	Colleague	0.3	Citizen	1.9	0.2
	0.2	Colleague	0.2	Colleague	←	0.2	Citizen	0.2	Colleague		
								0.1	Crane operator		
1.9	0.3	Citizen	0.3	Colleague	Reward	0.3	Colleague	0.4	Colleague	0.9	0.2
	0.3	Colleague	0.3	Citizen	→	0.2	Peer	0.2	Citizen		
						0.2	Citizen				
1.8	0.4	Citizen	0.4	Citizen	Reward	0.3	Citizen	0.3	Citizen	1.8	0.2
	0.3	Colleague	0.3	Colleague	←	0.2	Colleague	0.3	Colleague		
						0.2	Peer				

Table 2. MIS-ALIGNED agents: The distributions over denotative labels are cut off when the cumulative probability exceeds 0.5. d= deflection, f fundamental difference (smaller is better), Here the agents have parameters (Hank) $\alpha = 0.01$, $\beta = 0.001$, $\gamma = 0.25$ and (Tom) $\alpha = 0.5$, $\beta = 0.05$, $\gamma = 0.01$. There are 17 rounds in total with six rounds removed from the middle to make space. Tom and Hank are not able to come to an alignment here: although Hank sort of feels things are going ok, Tom definitely doesn't.

Hank											Tom	
	Tank		Tom		Behav	Tom		Hank				
d	Id	p	id	P	$\leftrightarrow$	Id	p	id	p	d	Δf
0.1	0.5	Citizen	0.5	Citizen	Test	0.5	Coal miner	0.5	Prosecutor	0.0	3.9
	0.5	Environmentalist	0.5	Culprit	$\rightarrow$	0.5	Coal miner	0.5	Prosecutor		
3.0	0.4	Environmentalist	0.6	Culprit	Test	1.0	Coal miner	1.0	Prosecutor	1.3	2.9
	0.2	Peer			$\leftarrow$						
2.4	0.2	Colleague	0.1	Culprit	Pacify	1.0	Coal miner	1.0	Prosecutor	4.0	3.4
			0.1		$\rightarrow$						
	0.1	Worker	0.1	Crane operator							
	0.1	Wage earner	0.1	Peer							
	0.1	Peer	0.1	Citizen							
				Liberal							
5.9	0.3	Peer	0.4	Culprit	Appease	1.0	Coal miner	1.0	Prosecutor	4.8	2.9
	0.2	Environmentalist	0.1	Malcontent	$\leftarrow$						
3.3	0.1	Citizen	0.1	Crane operator	Pacify	1.0	Coal miner	1.0	Prosecutor	6.7	3.3
	0.1	Peer	0.1	Malcontent	$\rightarrow$						
	0.1	Environmentalist	0.1	Culprit							
	0.1	Colleague	0.1	Foreman							
	0.1	Nonconformist	0.1	citizen							
6.1	0.3	Environmentalist	0.4	Malcontent	Pacify	1.0	Coal miner	1.0	Prosecutor	6.6	2.4
	0.2	Peer	0.2	Culprit	$\leftarrow$						

(continued)

Table 2. (continued)

Hank										Tom	
	Tank		Tom		Behav	Tom		Hank			
d	Id	p	id	P	↔	Id	p	id	p	d	Δf
2.1	0.2	Colleague			Test	1.0	Coal miner	1.0	Prosecutor	8.4	2.2
	0.1	Nonconformist	0.2	Citizen	→						
	0.1	Citizen	0.2	Crane operator							
	0.1	Foreman	0.1	Foreman							
2.3	0.2	Nonconformist		Crane operator	Obey	1.0	Miner	1.0	prosecutor	6.8	2.3
	0.1	Citizen	0.3	Foreman	←						
	0.1	union member	0.1	Citizen							
	0.1	crane operator	0.1								
2.2	0.2	Nonconformist	0.2	Crane operator	Obey	1.0	Coal miner	1.0	Prosecutor	5.3	2.0
	0.1	Crane operator	0.1	Foreman	←						
	0.1	Union member	0.1	Citizen							
	0.1	Foreman	0.1	Colleague							
2.2	0.2	Nonconformist	0.2	Crane operator	Discipline	1.0	Coal miner	1.0	Prosecutor	7.0	2.0
	0.1	Crane operator	0.1	Foreman	→						
	0.1	Union member	0.1	Citizen							
	0.1	Citizen	0.1	Colleague							
2.2	0.2	Nonconformist			Appease	1.0	Coal miner	1.0	Prosecutor	6.4	2.0
	0.1	Union member	0.2	Crane operator	←						
	0.1	Crane operator	0.2	Foreman							
	0.1	Citizen	0.1	Citizen							

Discussion

We presented an analysis of three different kinds of social uncertainty and the necessary tradeoffs between them as predicted by Bayesian affect control theory (Hoey et al., 2016, 2021; Schröder et al., 2016). A theoretical goal of this analysis was to show that some aspects of the dynamics of social orders are an emergent outcome of the affective management of uncertainty, beyond affective-meaning structures and impression dynamics studied so far by affect control theorists. A practical goal of our analysis was to show how BayesACT can serve as an analytical tool to better understand societies' adaptive responses to important global challenges such as ecological and technological change accompanied by social conflict. Of course, this contribution is purely theoretical and, in many ways, conjectural, so this paper is more the starting point of a research agenda.

The obvious next step will be to conduct systematic computational experiments to gain more confidence about the impact of uncertainty-parameter alignment on interaction dynamics. This is a daunting task given that the space spanned by the BayesACT parameters is theoretically very large. The brief simulations shown here as an example are enough to illustrate and provide some plausibility to our claims about the importance of uncertainty management in affective control of social interaction, especially in the context of significant societal change. However, we fully realize they are too specific to count as substantial evidence for our novel ideas put forth in this paper.

The next step would be to build simulation experiments with a BayesACT model of group interactions, where the tradeoffs between uncertainty parameters explored in this paper are systematically varied in the context of social systems beyond simple dyads. To enhance the argument that social structures have a functionality with regard to the types of problems a group needs to solve (e.g., Ridgeway, 2019), additional simulation experiments should explore which configurations of the three uncertainty parameters allow BayesACT agents in a group setting to achieve socially beneficial outcomes.

Of course, testing the proposed uncertainty-management mechanisms will also require empirical work. Partly, this will involve validation of BayesACT simulations. For example, choices of participants in experimental social dilemma situations can be compared with simulation output, or actual participants can play games against artificial agents run by BayesACT mechanisms. Moreover, ACT-style survey work can serve to test some of the assumptions outlined here. Questions such as the following can be studied with existing techniques: Is the consensus about affective meanings of identities relevant to policy challenges becoming smaller over time? Are specific affective meanings becoming more associated with political views? Other questions will require the development of novel empirical measures. For example, if political views *do* reflect different preferences for affective uncertainty management (coherence vs. dependence precision), identifiable contestants in relevant debates will exhibit different styles of uncertainty management. Testing this hypothesis will require the development of empirical measures of the three uncertainty parameters in BayesACT.

Lastly, uncertainty management at the connotative level is a key factor in practical issues of change and transition management. Climate change communication might be much less controversial if it entails positive visions of a future social order where many of the connotative status positions of today can be preserved. In contrast, the doomsday scenarios often invoked by environmental activists not only have the desired effect of creating a sense of urgency in the political arena, but also cause powerful resistance at the connotative level as a result of a threatened sense of identity. Such reactions cannot be overcome by more objective information about the problem, but require increased certainty about the connotative status affected people will be able to find in the emerging society of the future.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work benefited from funding through the Transatlantic Partnership Digging Into Data Platform (NSERC grant #5660-501727 and SSHRC grant #869-2016-0005 to Hoey and DFG grant #SCHR1282/3-1 to Schröder).

Data Availability

See bayesact.ca

ORCID iD

Tobias Schröder https://orcid.org/0000-0002-7113-7464
Jesse Hoey https://orcid.org/0000-0001-5340-2204

Notes

1. All data in this paper from the USA 2015 dataset (Smith-Lovin, et al., 2016).
2. We use *(in)validity* here in the sense defined by Kahneman and Klein (2009) as how predictable or unambiguous the future is according to the model, not in the psychometric sense.
3. In the mathematical BayesACT model (Hoey et al., 2021), the parameters δ, α, and γ are real positive numbers that refer to the *dispersion* of the distributions (the *variance* if the distribution is normal). Here, we are describing parameters as the *precision* of the distribution, and we write $\widehat{\boldsymbol{\delta}},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\gamma}}$ (which is the same as $\delta^{-1},\alpha^{-1},\gamma^{-1}$). In the results section, we revert to the un-decorated parameters.
4. A fourth parameter, β, represents connotative dynamics, but can be set in proportion to α, and so we leave it from the subsequent discussion.

5. While starting with very dispersed distributions can lead to a convergence if the dynamics and control mechanisms are precisely shared amongst group members, if all three are dispersed, the task becomes much more difficult.
6. The second author owes this example to a discussion with the mayor of Spremberg, a community in East Germany heavily affected by Germany's decision to phase out lignite mining by 2038 at the latest.
7. See the supplementary page at bayesact.ca with code and data for full list of identities and behaviors.
8. This particular setting was found through a grid search in the parameter space as the best to find aligned solutions with low deflection. The two parameter sets may not always be identical because of the inherent problem of "turn taking": one actor must act first in any simulation, and the choice of who goes first makes a difference in how the parameters that best fit go together. For example, a more seamless interaction would result if the person going first is given an identity that is more powerful than the person going second.
9. That is, if a/c are the identity distributions for actor (a) and object (c) for agent 1, and b/d are the identity distributions of agent 2, then this is $(a\text{-}d)^2+(c\text{-}b)^2$.
10. see footnote 3 - we revert to parameters denoting dispersion here (un-hatted symbols), rather than precision (hatted symbols)

References

Ambrasat, J., von Scheve, C., Conrad, M., Schauenburg, G., & Schröder, T. (2014). Consensus and stratification in the affective meaning of human sociality. *Proceedings of the National Academy of Sciences*, *111*(22), 8001–8006. https://doi.org/10.1073/pnas.1313321111.

Åström, K. J. (1965). Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, *10*(1), 174–205. https://doi.org/10.1016/0022-247X(69)90163-2.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Putnam's sons.

FeldmanHall, O. & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature Human Behaviour*, *3*(5), 426–435. https://doi.org/10.1038/s41562-019-0590-x.

Freeland, R. E., & Hoey, J. (2018). The structure of deference: Modeling occupational status using affect control theory. *American Sociological Review*, *83*(2), 243–277. https://doi.org/10.1177/0003122418761857.

Fukuyama, F. (2018). *Identity: Contemporary identity politics and the struggle for recognition*. Profile Books.

Gilead, M., Trope, Y., & Liberman, N. (2020). Above and beyond the concrete: The diverse representational substrates of the predictive brain. *Behavioral and Brain Sciences*, *43*(e121), 1-74. https://doi.org/10.1017/S0140525X19002000.

Heise, D. R. (2007). *Expressive order: Confirming sentiments in social actions*. Springer.

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review*, *119*(2), 304–320. https://doi.org/10.1037/a0026767.

Hoey, J. (2021). Freedom and equality as uncertainty in groups. *Entropy*, *23*(11), 1384. https://doi.org/10.3390/e23111384.

Hoey, J., MacKinnon, N., & Schröder, T. (2021). Denotative and connotative management of uncertainty: A computational dual-process model. *Judgement and Decision Making*, *16*(2), 505505–505550. https://EconPapers.repec.org/RePEc:jdm:journl:v:16:y:2021:i:2.

Hoey, J., Schröder, T., & Alhothali, A. (2016). Affect control processes: Intelligent affective interaction using a partially observable Markov decision process. *Artificial Intelligence*, *230*(January), 134-172. https://doi.org/10.1016/j.artint.2015.09.004.

Hoey, J., Schröder, T., Morgan, J., Rogers, K. B., Rishi, D., & Nagappan, M. (2018). Artificial intelligence and social simulation: Studying group dynamics on a massive scale. *Small Group Research*, *49*(6), 647–683. https://doi.org/10.1177/1046496418802362.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755.

MacKinnon, N. J., & Heise, D. R. (2010). *Self, identity, and social institutions*. Palgrave Macmillan.

MacKinnon, N. J., & Hoey, J. (2021). Operationalizing the relation between affect and cognition with the somatic transform. *Emotion Review*, *13*(3), 245–256. https://doi.org/10.1177/17540739211014946.

Morgan, J. H., Zhao, J., Zöller, N., Sedlacek, A., Chen, L., Piper, H., Beck, Y., Rogers, K. B., Hoey, J., & Schröder, T. (2021). Modeling the culture of online collaborative groups with affect control theory. In P. Ahrweiler & M. Neumann (Eds.), *Advances in Social Simulation* (pp. 147–160). Springer. https://doi.org/10.1007/978-3-030-61503-1_14.

Osgood, C. E. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, *12*(3), 194–199. https://doi.org/10.1037/h0027715.

Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress*. Penguin.

Ridgeway, C. L. (2019). *Status: Why is it everywhere? Why does it matter?* Russell Sage Foundation.

Rockström, J., Steffen, W., Noone, K., Persson, A., Chapin, F. S. III, Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., …& Foley, J. A. (2009). Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and Society*, *461*(2), 472–475. https://doi.org/http://www.ecologyandsociety.org/vol14/iss2/art32/.

Schneider, A. (2006). *Mean affective ratings of 787 concepts by Texas tech university undergraduates in 1998 [computer file]*. Program Interact. Distributed at Affect Control Theory Website http://www.indiana.edu/~socpsy/ACT/interact/JavaInteract.html.

Schröder, T., Hoey, J., & Rogers, K. B. (2016). Modeling dynamic identities and uncertainty in social interactions: Bayesian affect control theory. *American Sociological Review*, *81*(4), 828–855. https://doi.org/10.1177/0003122416650963.

Smith-Lovin, L., Robinson, D. T., Cannon, B. C., Clark, J. K., Freeland, R., Morgan, J. H., & Rogers, K. B. (2016). *Mean affective ratings of 929 identities, 814 behaviors, and 660 modifiers by University of Georgia and Duke University undergraduates and by community members in Durham, NC, in 2012-2014*. University of Georgia. Distributed at UGA Affect Control Theory Website. http://research.franklin.uga.edu/act/.

Vaisey, S. (2009). Motivation and justification: A dual-process model of culture in action. *American Journal of Sociology*, *114*(6), 1675–1715. https://doi.org/10.1086/597179.

Vaisey, S. & Valentino, L. (2018). Culture and choice: Toward integrating cultural sociology with the judgment and decision-making sciences. *Poetics*, *68*, 131-143. https://doi.org/10.1016/j.poetic.2018.03.002.

WBGU (2019). *Towards our common digital future*. Flagship report. German Advisory Council on Global Change. https://www.wbgu.de/en/publications/publication/towards-our-common-digital-future.

Zöller, N., Morgan, J. H., & Schröder, T. (2020). A topology of groups: What GitHub can tell us about online collaboration. *Technological Forecasting and Social Change*, *161*, 120291. https://doi.org/10.1016/j.techfore.2020.120291.

Author Biographies

Jesse Hoey is a professor in the David R. Cheriton School of Computer Science at the University of Waterloo, where he leads the Computational Health Informatics Laboratory. He holds a PhD degree (2004) in computer science from the University of British Columbia.

Tobias Schröder is a professor in the Department of Social and Educational Sciences at the Potsdam University of Applied Sciences, where he currently serves as the Vice President for Research and Innovation. He holds a doctoral degree (2009) in psychology from the Humboldt University in Berlin, Germany.