PHILOSOPHICAL PSYCHOLOGY
https://doi.org/10.1080/09515089.2023.2224379

Routledge
Taylor & Francis Group

Guess who? Identity attribution as Bayesian inference

Francesco Rigoli

Department of Psychology, University of London, London, UK

ABSTRACT

An influential argument is that mental processes can be explained at three different levels of analysis: the functional, algorithmic, and implementation level. Identity attribution (the process whereby an identity is attributed to another individual or to the self) has been rarely explored at the functional level. To address this, here I propose a theory of identity attribution grounded on Bayesian inference, being the latter a well-established functional perspective in cognitive science. The theory posits that an identity is inferred based on observations about a target's features, about the context, and about motivational factors. This inference can be made based upon multiple sources of observations, with prior beliefs becoming more prominent when observations are fewer in number. The theory offers an interpretation of key processes driving identity attribution, potentially providing a platform for integrating different perspectives on identity in psychology and sociology.

ARTICLE HISTORY
Received 19 December 2022
Accepted 31 May 2023

KEYWORDS
Identity; identity attribution; Bayesian; context; computational; inference

1. Introduction

An identity describes a social category attributed to another individual or to the self (in the latter case, referred to as self-identity) (Burke & Stets, 2009; Hogg, 2020; Weigert et al., 2007). Attributing an identity means classifying a person as belonging to a specific group (e.g., to a specific profession such as carpenter) and not to other groups (e.g., other professions such as teacher, peasant, or shopkeeper), thereby evoking a set of specific norms and expectations (e.g., about a person's characteristics and behavior) (Terry et al., 1999). Moreover, by eliciting specific affective and behavioral tendencies, attribution of an identity has immediate practical implications (Burke & Tully, 1977; Callero, 1985). For instance, someone identified as a professional athlete might be expected to be energetic, and in turn this expectation might predispose one toward enacting an energetic behavior when interacting with the person.

CONTACT Francesco Rigoli francesco.rigoli@city.ac.uk Department of Psychology, City University of London, Northampton Square, London EC1V 0HB, UK

The notion of identity is at the crucible of psychological and sociological levels of analysis: although ultimately an identity is attributed by an individual mind, it is the social context which determines which identities are salient (Weigert et al., 2007). In turn, identity attribution drives people's behavior in social contexts: for instance, when a specific identity is attached to oneself (e.g., as a professional athlete), one's own behavior will conform to that identity (e.g., one will behave more energetically), thereby shaping social interactions (Brewer, 1991; Hogg, 2004). Likewise, the way we approach others depend on which identity we attribute to them (Tyler & Blader, 2001). In short, identities shape, and at the same time are shaped by, social dynamics.

Given the centrality of identity attribution to both psychological and sociological domains, explaining this process is highly important. An influential argument (Marr & Poggio, 1976) posits that any mental process can be explained at three different levels: a functional level (analyzing the general function of the phenomenon), an algorithmic level (focusing on the specific cognitive processes underlying the phenomenon), and an implementation level (exploring how these processes unfold in the brain). A rich theoretical literature has examined identity attribution at an algorithmic level (e.g., Burke & Stets, 2009; Hogg, 2020; Weigert et al., 2007), greatly advancing the scientific understanding of the process. Yet, a problem of algorithmic theories is that the principles they advocate are not grounded on identifying what the process accomplishes, and thus they may appear somewhat ad-hoc. Nonetheless, the functional level, which is exatly the one focusing on what the process accomplishes, remains to be explored in relation with identity attribution. This raises the question of whether adopting a functional approach can offer any new insight that goes beyond existing theories. In psychology, Bayesian inference is amongst the most influential approaches to investigate the functional level (Clark, 2013; Hogg et al., 1995; Oaksford & Chater, 2007). The key idea of this approach is that fundamental cognitive processes such as perception (Kersten et al., 2004; Pettigrew & Meertens, 1995), memory (Hemmer & Steyvers, 2009; B. M. Turner et al., 2013), and decision-making (Botvinick & Toussaint, 2012; Friston et al., 2015) can be interpreted as expressions of Bayesian inference, where prior beliefs are integrated with novel information according to Bayesian rules. Can Bayesian inference help understanding identity attribution? Besides shedding light on identity attribution itself, addressing this question can potentially highlight the links between identity attribution and other cognitive processes for which a Bayesian interpretation is well-established, thus contributing to build a unifying picture of how the mind works.

Research employing a Bayesian outlook in order to examine the topic of identity is rare. A first body of work has adopted Bayesian principles

(specifically, the notion of structure learning) to study how social categories are learnt based on past experience (Gershman & Cikara, 2020; Lau et al., 2018). Yet, this investigation has focused exclusively on learning, that is, on how experience shapes the formation of social categories. The issue of how social categories, once acquired, are next employed to infer an identity (i.e., the issue of identity attribution) has been neglected by this research. Another relevant body of work (Schröder et al., 2016) has recently developed a Bayesian framework to enrich Affect Control Theory (Heise, 1987), which is one of the most influential sociological perspectives on how social interactions work. Like its predecessor, Bayesian Affect Control Theory is very broad in scope, examining the interaction among various processes such as impression formation, affective appraisal, communicative behavior, and, of interest here, identity processing. The advantage of such open-ended approach is that it acknowledges the complex interactions occurring among the multiple psychosocial processes engaged during social interactions. Yet, the drawback is that this approach is not ideal for isolating the basic mechanisms at play specifically during identity attribution. These mechanisms, thus, remain to be elucidated adopting a Bayesian outlook.

To fill this gap in the literature, the present paper examines whether the tools of Bayesian inference can offer any insight about the specific processes underlying identity attribution. The focus is on the typical scenario considered by previous literature on the topic (Burke & Stets, 2009; Hogg, 2020; Weigert et al., 2007): a scenario where an agent needs to select an identity based on a set of categorical variables. The following section overviews my proposal, which is referred to as Bayesian Inference Model of Identity (BIMI). Next, to clarify the functioning of the model, this is assessed in a variety of simulations. Finally, BIMI is appraised with respect to previous literature on identity attribution.

2. Bayesian inference model of identity (BIMI)

BIMI describes the processes whereby a person attributes one specific identity category to a target individual (the target being the self or someone else). Flexibility is at the heart of identity attribution: depending on the context, the same person can be identified as a doctor, as a woman, or as an African-American (J. C. Turner et al., 1987). Thus, understanding identity attribution requires explaining why, in a certain context, a specific identity arises instead of alternatives. BIMI interprets this by proposing that, based on some available information, an identity is attributed following a Bayesian inference process (Rigoli et al., 2017). A useful way to overview the model is to rely upon the Bayesian Network formalism (Figure 1)[1] (Bishop, 2006). Here, the brain is assumed to represent (i) a set of important variables (in BIMI, all categorical),

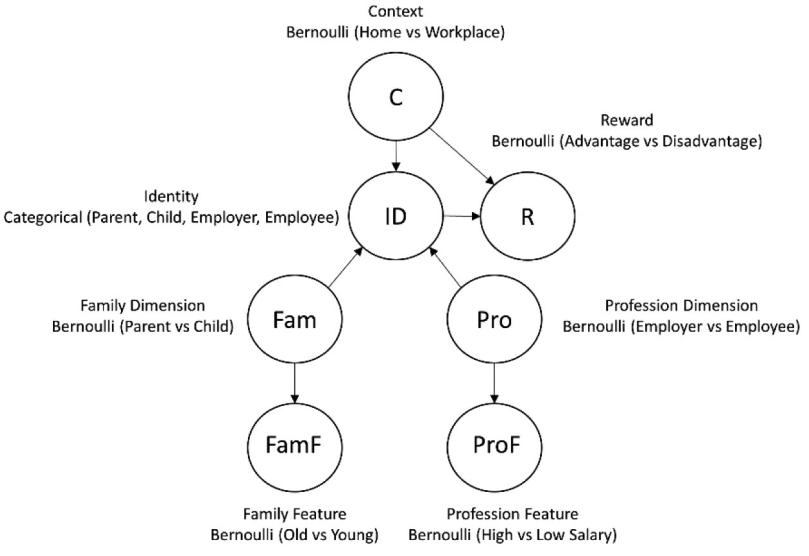

Figure 1. Graphical model proposed by BIMI. Variables (represented by circles) include Context (C), Identity (ID), Reward (R), Family (Fam), Profession (Pro), Family Features (FamF), and Profession Features (ProF). Arrows indicate probabilistic dependencies among variables.

described graphically by circles, and (ii) their probabilistic dependencies, described by arrows. At the center of the Network is the variable Identity (ID), representing a list of potential identities that can be attributed. This list adds up multiple identity dimensions. Consider an example where two dimensions are envisaged, including the profession dimension (having employee and employer as categories) and the family dimension (having parent and child as categories). Since ID encompasses all categories across dimensions, here it comprises four categories, including employee, employer, parent, and child. The value of ID reflects the category (e.g., employer) in the list which is appropriate for describing a target individual in a specific situation. However, BIMI assumes that ID is a *latent* (or *hidden*) variable; in other words, that it cannot be directly observed, but needs to be inferred from other information. Intuitively, this captures the fact that one does not know with certainty which identity is currently the most appropriate, though one can come up with an informed guess. Inferring ID is the key process realized by BIMI.

Context (C), reflecting surrounding contextual information, is the second variable in the model. For example, C might include two categories: workplace versus home. As the arrow going from C to ID indicates, BIMI assumes that the former variable influences the second; in other words, it assumes that which identity is salient depends on the ongoing context. For example, BIMI can capture the fact that the categories of ID linked with the profession dimension (employee and employer) are more likely to be salient

in the workplace, while the categories of ID linked with the family dimension (parent and child) are more likely to be salient at home. BIMI assumes that the variable C can be directly observed.

Besides contextual information, assessing the characteristics of a target individual (e.g., asking whether someone is a parent, child, employer, or employee) is obviously critical to infer her appropriate ID. To account for this, BIMI includes variables describing each identity dimension separately, in the example corresponding to Profession (Pro; having employee versus employer as categories) and Family (Fam; having parent versus child as categories). Both variables project to ID, capturing the reasoning that, in conjunction with the ongoing context, the target's professional and family role is what matters for inferring her identity. Both Pro and Fam are treated as latent variables, as these cannot be known directly but need to be inferred from more basic features. These features are captured by separate variables, in the example including Profession Features (ProF; having high salary versus low salary as categories) and Family Features (FamF; having old age versus young age as categories). Feature variables are treated as manifestations of underlying dimension variables, as the arrows going from Fam to FamF and from Pro to ProF, respectively, indicate (reflecting the belief that, for instance, a parent will be more likely to be old compared to a child). Feature variables can be observed directly and used to infer ID. For the sake of simplicity, here I include one feature variable per dimension; however, in principle multiple feature variables can be considered for any single dimension. For instance, feature variables associated with the profession dimension (Pro) might include salary (high versus low), years of employment (many versus few), and education (short versus long).

One last variable is Reward (R), capturing the expected consequences (in terms of reward or punishment) associated with attributing a specific identity ID to the target individual in a specific context C. Thus, R depends on both ID and C. For the sake of simplicity, R can assume one of two values: advantage (indicating that a positive outcome is expected) or disadvantage (indicating that a negative outcome is expected). Intuitively, R indicates how likely advantages or disadvantages are when a specific identity (e.g., employer) is selected in a specific context (e.g., in the workplace). For example, the variable R might indicate that a child self-identity in the workplace is expected to produce disadvantages more often than advantages. Below, we will see how the inclusion of the variable R accounts for the role of motivation in identity attribution.

In summary, BIMI comprises variables critical for inferring the identity of a target individual in a specific context. Though such identity is not directly observable, the argument is that this can be inferred from two

potential sources of information: the ongoing context and the individual's features regarding a set of dimensions. Moreover, the model proposes that this inference can be influenced by motivational dynamics as captured by the reward variable. Below, I will examine implications of BIMI by assessing the model in different simulations.[2]

3. How BIMI works

3.1. Inference with all information available

First, let us explore cases where all potentially observable variables are available during inference. These variables include C, FamF, and ProF (R can also be treated as observed, as we shall see below). In this scenario, inference corresponds to estimating the posterior probability distribution P (ID | C, FamF, ProF). Here, each identity is associated with a posterior probability or *weight* describing how salient that identity is in a certain situation. I assume that the final identity attributed to the target person corresponds to the identity having the highest weight (a different rule could be used here: the attributed identity could be randomly chosen from a distribution where, for each identity, the probability to be chosen is equal to its weight; R. D. Luce, 1959). Moreover, from the posterior probability P(ID | C, FamF, ProF) the level of uncertainty U about identities can be derived.[3] This is null when one identity has a weight equal to one (and all other identities have a weight equal to zero; i.e., when there is absolute certainty about the appropriate identity) and it is maximal when all identities have equal weight (i.e., when all identities look exactly equally salient). The notion of uncertainty might offer insight on important psychological aspects of identity attribution. For example, low uncertainty might be at play when people consider identities as naturally given, thus failing to acknowledge that identities are in fact socially constructed and thus always to some degree uncertain (Reicher, 2004). Moreover, the notion of uncertainty might be relevant to interpret the phenomenon of identity interference. Empirical research reveals that, when multiple identities are salient, different identities can be simultaneously evoked and interfere with one another at the behavioral level (Oyserman et al., 2006; Settles, 2004). BIMI explains identity interference as emerging when uncertainty is high, in other words when multiple identities have similar weight.

Figure 2 displays the posterior probability P(ID | C, FamF, ProF) for the four ID categories in the example (employer, employee, parent, child), considering different values of C, FamF, and ProF. First, results show that the variable C determines whether either FamF or ProF count. For instance, a target associated with FamF = old and ProF = high salary is identified as employer when C = workplace and as parent when C = home. Second, when

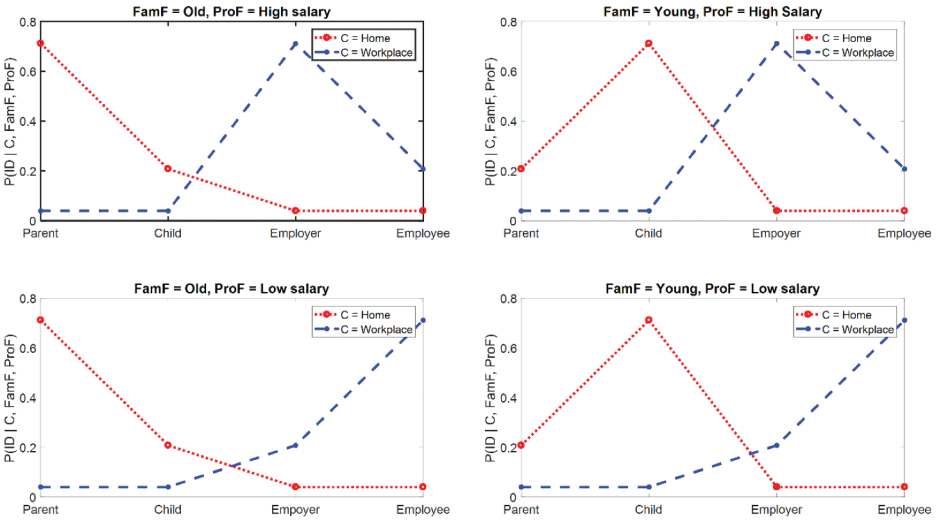

Figure 2. Role of observations. Posterior of the variable ID (P(ID | C, FamF, ProF)) for different values of C, FamF, and ProF.

the value of C is kept constant, variations in the salient dimension (i.e., changes in FamF when C = home, and changes in ProF when C = workplace) matter, while variations in the irrelevant dimension do not: when C = home, a target person is identified as parent when FamF = old and as child when FamF = young, independent of the value of ProF. Likewise, when C = workplace, a target person is identified as employer when ProF = high salary and as employee when ProF = low salary, independent of the value of FamF.

Overall, this simulation highlights a different role for the context and for the feature variables. The context is not determinant as such, but it sets the stage by establishing which dimension becomes critical. For example, the context prescribes that in the workplace the critical dimension becomes the profession, which, based on salary, distinguishes employers from employees. In this context, the family dimension, distinguishing parents from children based on age, should be ignored.

3.2. *Inference without knowing the context*

Not in all cases contextual information is available, and yet within BIMI an identity can be inferred also in these cases. An instance of this is when an agent encounters a new context and has no idea of which identity dimensions matter in this context. In the example above, the agent is familiar with two contexts only: the workplace and home. What happens when, say, the agent finds herself in a restaurant? Will she categorize other people according to the family dimension (i.e., as children or parents) or according to the profession dimension (i.e., as employees or employers)?

BIMI argues that, when FamF and ProF are observed, these can be relied upon to infer ID even in the absence of any contextual information C. Formally, this inference requires estimating the posterior probability P(ID | FamF, ProF) (the uncertainty U can be derived here too - see note 2 , but now removing C from the observed variables). Because the context is not observed, the context prior probability P(C) now becomes critical. This reflects beliefs about how likely the home and the workplace contexts are to occur a priori. Intuitively, when at the restaurant, P(C) indicates how likely it is that the restaurant works as the workplace (implying that the profession is the relevant dimension) and how likely it is that the restaurant works as home (implying that the family is the relevant dimension). Figure 3 illustrates how variations in P(C) affect the posterior probability P(ID | FamF, ProF) for a target associated with FamF = young and ProF = low salary. As P(C = home) increases (i.e., as one increasingly believes that the home context is more likely a priori), the inferred identity moves progressively from employee to child.

This scenario offers insight on important psychological phenomena. First, it explains why humans often exhibit a tendency to categorize people according to specific criteria, for example based on age, social class, or ethnic group (J. C. Turner et al., 1987). BIMI interprets this tendency as due to high prior probability for the context associated with any of these criteria. In turn, the prior probability for the context depends on past experience, namely on how common each context has been in the past. For example, a predisposition to categorize people based on their social class might result from living within a community where people are usually judged based on this criterion. A tendency to categorize people in a certain way emerges when the context is novel, and is expressed by

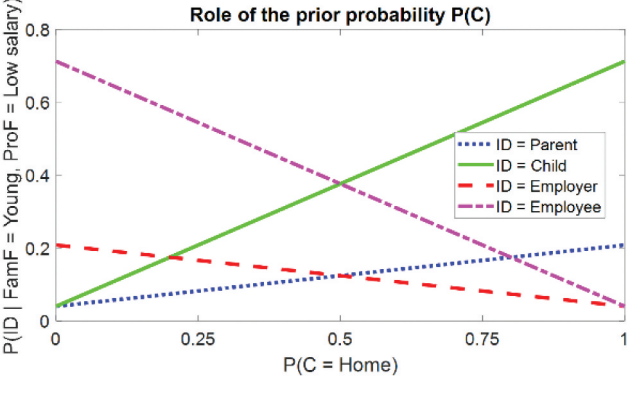

Figure 3. Role of the prior probability of the context. Posterior of the variable ID (P(ID | FamF = Young, ProF = Low salary)) for different prior probability of the variable C (P(C)).

generalizing a prevailing context toward novel contexts. Moreover, this tendency might also arise when contextual information is potentially available but it is disregarded.

One last phenomenon that can be described by this scenario concerns self-identities, namely circumstances where one is attributing an identity to oneself. At any given moment, a person can incarnate one among several potential self-identities, such as mother, physician, or African-American. Sometimes, which self-identity is selected depends on considering contextual information (e.g., the mother self-identity will be activated with children, the physician self-identity at work, and the African-American self-identity when voting). However, when contextual information is absent or disregarded, a default self-identity will be enacted (as particularly emphasized by the Identity Theory research tradition; Stets & Burke, 2000; Stryker & Serpe, 1982). BIMI explains default self-identities as arising from high prior probability attributed to a specific context. For example, attributing high prior probability to the workplace context will result in a predisposition to embody the role of physician, even at home or when voting.

3.3. Inference without knowing the target

Sometimes, although information about context is available, information about the target person to identify is poor. In our example, this is reflected by lack of information about FamF and ProF. All the same, BIMI argues that an identity can still be inferred, now by estimating the posterior probability P(ID | C) (the uncertainty U can be derived here too - see note 2, but now removing FamF and ProF from the observed variables). In this case, prior beliefs about FamF and ProF (i.e., P(Fam) and P(Pro), respectively), capturing the perceived likelihood of the categories of each dimension, become even more critical. Considering Fam as an example, P(Fam) describes the perceived prior likelihood of meeting parents and children, respectively. Figure 4 illustrates the posterior probability P(ID | C) for different contexts and when varying either P(Fam) or P(Pro). When C = home, variations in P(Pro) are uninfluential, while, as the prior probability P(Fam = parent) increases, the inferred identity moves progressively from child to parent. Likewise, when C = workplace, variations in P(Fam) are uninfluential, while, as the prior probability P(Pro = employer) increases, the inferred identity moves progressively from employee to employer.

Psychologically, this scenario accounts for a tendency to attach default identities in certain contexts, for example a tendency to label people as fellow citizens (instead of foreigners), or as poor (instead of rich), or as criminal (instead of honest) in certain contexts. This tendency is interpreted by BIMI as arising from a high prior probability associated with the default

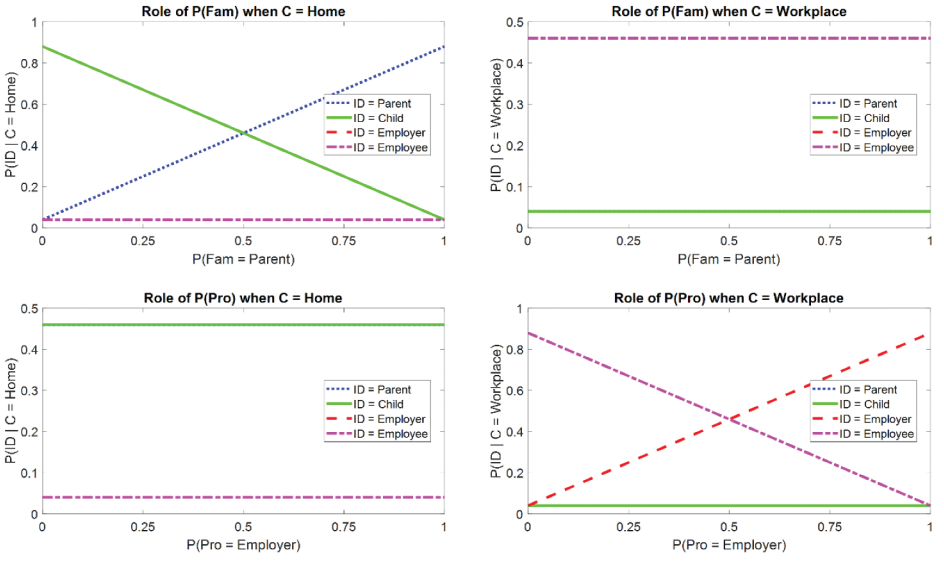

Figure 4. Role of the prior probability of identity dimensions. Posterior of the variable ID (P(ID | C)) for different prior probability of the variable Fam (P(Fam); on the top) or of the variable Pro (P(Pro); on the bottom). Note that some lines are not visible because they overlap with other lines.

identity. A default identity might come into play when a new person is encountered and no information about her is available (i.e., when feature variables are unknown). A default identity might also be relied upon when relevant information about the person is disregarded despite being potentially available (e.g., when someone is labeled as a criminal a priori, without really assessing whether this label is appropriate) (Dovidio et al., 1997). A default identity can be developed from experiencing a prevalence of exemplars belonging to that identity (e.g., a tendency to view people as fellow citizens might simply derive from meeting fellow citizens more often than foreigners), but it might also derive from other sources of social influence (e.g., rumors or media), potentially expressing poorly grounded forms of prejudice (Bissell & Parrott, 2013). As this consideration hints, BIMI offers a formal perspective about the nature of prejudice as well as of stereotype; this point will be examined in the next section.

3.4. Stereotype and prejudice

BIMI can offer a new perspective to look at the nature of prejudice (Duckitt, 1992) and stereotype (Fiske, 1998). To see how prejudice is interpreted by BIMI, the key point is that (as just examined) sometimes an identity is inferred even when some potential information is absent (or disregarded). When inference relies on partial, and not full, information, we can talk

about prejudice. This fits with Allport's classical definition of prejudice as "a feeling, favorable or unfavorable, toward a person or thing prior to, or not based on, actual experience" (Allport, 1954).

Above, we have already examined cases of prejudice, considering scenarios where an identity is inferred in the absence of information about context or about feature variables. Let us now consider a slightly more complex scenario in which the profession dimension is associated with three feature variables: Salary (SAL, with high versus low salary as categories), Years of employment (EMP, with many versus few years as categories), and Education (ED, with short versus long education as categories). In this scenario, multiple sources of information can be considered to infer an identity: FamF, C, SAL, EMP, and ED. Imagine that all this information is available except for ED. Based on the available information, the posterior probability P(ID | C, FamF, SAL, EMP) can be estimated. This inference represents a form of prejudice because it is performed without considering all potentially relevant information (in this case without considering ED). This logic can be applied to interpret forms of prejudice at the origin of discrimination in society. For example, just knowing that someone comes from a specific country (or neighborhood, or ethnic group) might be considered enough to categorize the person as violent, stupid, or ignorant, without assessing other features that are in fact essential for a correct categorization.

The same scenario can be considered to assess the notion of stereotype, defined as a generalized belief about the features possessed by an individual belonging to a specific category (Fiske, 1998). Not only the model can be used to infer ID, but it can also be employed to infer ED (remember that this variable is not observed, and thus unknown) by calculating the probability distribution P(ED | C, FamF, SAL, EMP) (McCauley et al., 1980). In other words, within BIMI specific features can be inferred even when they are not observed. This form of inference fits with the classic definition of stereotype, as it requires applying beliefs about the general identity to predict a single feature. This logic can be used to interpret forms of stereotype at the origin of discrimination in society. For example, after one is categorized as a woman, or black, or poor, an inference about her features (e.g., intelligence) might follow (e.g., the person might be expected to have poor intelligence), even though these features are actually unknown. Note that the proposed difference between prejudice and stereotype is subtle: the former involves inference of an identity, which is a latent variable, whereas the latter involves inference of a feature, which is a variable that can potentially be observed.

The interpretation of prejudice and stereotype proposed here might appear as problematic at first: when making judgments, we rarely consider all potentially available information, implying that BIMI would regard most

everyday-life inferences as instances of prejudice or stereotype – a conclusion which appears as unwarranted. However, this criticism can be addressed by viewing prejudice and stereotype not as qualitatively different from other forms of inference, but as extreme cases that violate social norms concerning how much information should be considered during inference. In other words, while most everyday-life inferences would disregard an acceptable amount of information (acceptable in terms of social norms), sometimes people's inferences would disregard an amount of information which is inacceptable, resulting in full-fledged prejudice or stereotype.[4]

In summary, BIMI offers a formal description of both prejudice and stereotype. The former is interpreted as arising from inferring an identity from incomplete information, while the latter is explained as arising from inferring a feature (or a set of features) based on available information.

3.5. The role of reward

Let us now examine the role of the reward variable R. Consider our example above (where possible identities are child, parent, employer, and employee), but now adding that employers, when encountered, require to be treated with particular respect (otherwise, say, one might risk losing the job). Thus, in this scenario, failing to identify an employer is somewhat more risky than failing to identify an employee, or parent, or child. These and similar incentives are captured by BIMI by setting a higher probability to an advantage every time a target is identified as an employer (compared to when the target is identified as someone else). Formally, this is implemented by setting P(R = advantage | C, ID = employer) = 1 and at the same time by setting P(R = advantage | C, ID = employee) = 0.5, P(R = advantage | C, ID = child) = 0.5, and P(R = advantage | C, ID = parent) = 0.5. Once this is implemented, the influence of R emerges by treating this variable as observed and by setting it to advantage during inference of ID. The left panels of Figure 5 display the posterior probability P(ID | R = advantage) (note that the context and feature variables are unknown in this instance; however the same effect would emerge if feature variables were known), in which R is treated as observed and is set to advantage. The figure considers different values of the prior probability P(C) and P(Pro). Although P(C) and P(Pro) remain important, the results show a consistent bias in favor of identifying people as employers. For comparison, the right panels of Figure 5 explore the same scenario, but now estimating the marginal probability P(ID) without considering R as observed. Now, the bias in favor of identifying people as employers has disappeared.

Implementing the variable R enables BIMI to explain why identity attribution is not an impartial process, but, to some degree, it aims at fulfilling

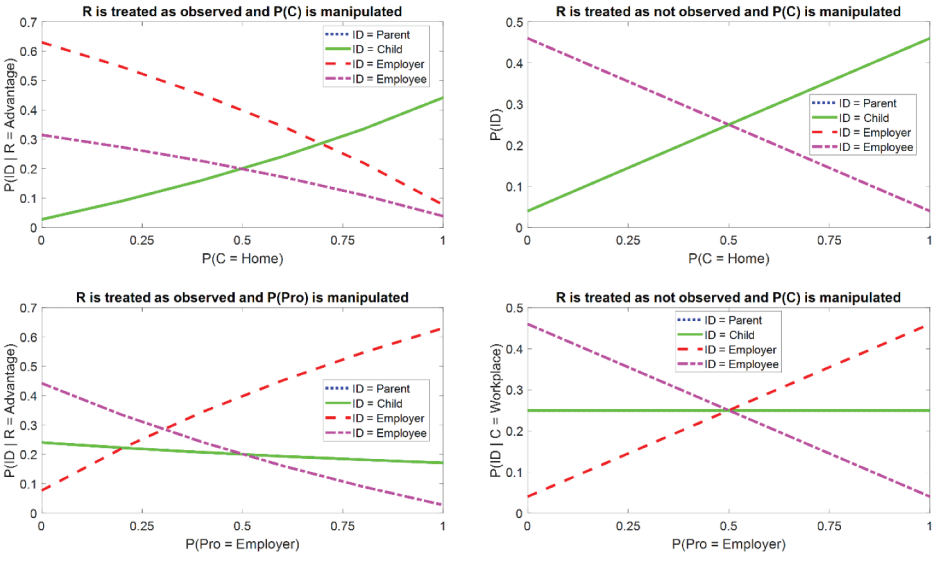

Figure 5. Role of the reward variable. Left panels display the Posterior of the variable ID (P(ID | R = Advantage)) for different prior probability of the variable C (P(C); on the top) or of the variable Pro (P(Pro); on the bottom). Right panels display the marginal of the variable ID (P(ID); now the reward variable R is not treated as observed) for different prior probability of the variable C (P(C); on the top) or of the variable Pro (P(Pro); on the bottom). Note that some lines are not visible because they overlap with other lines.

motives such as satisfying self- or group-interests (Bobo & Kluegel, 1993; Fiske, 1998; Pettigrew & Meertens, 1995). Evidence suggests that sometimes we categorize others in a way which is more convenient for ourselves (e.g., Bobo & Kluegel, 1993; Pettigrew & Meertens, 1995). For example, in a competitive environment, people tend to attach a more negative label to others (Suls et al., 2013; J. C. Turner, 1975). Self-identity is not immune to similar effects, as highlighted by evidence indicating that we tend to embody the identity which appears more favorable in a given circumstance (Burke & Reitzes, 1991; Stryker & Burke, 2000). BIMI captures these motivational factors by including the reward variable R and assuming that advantages will be experienced, thus (other things being equal) leading to inferring the identity associated with better prospects. Notably, as also evidence indicates (Rigoli, 2021), in BIMI motivational factors are not totally unconstrained, but are limited by other factors such as context and feature variables (as Figure 5 illustrates).

3.6. Dependent and hierarchical dimensions

So far, we have assumed that dimensions (i.e., Pro and Fam) are independent. This assumption implies that knowing one's profession is irrelevant for knowing one's family role, and vice versa. This assumption can be

modified in such a way that now one dimension is assumed to influence the other (McGarty, 1999). In our example, this can be implemented in the graphical model by adding an arrow going from Pro to Fam. This captures the belief that, if one is an employer, then the same person will be more likely to be a father. Situations where dimensions are dependent are rife in the real world: country and religion, gender and profession, ethnic group and language, are all examples of intermingled dimensions. Similar to Figures 2, Figure 6 describes the posterior probability P(ID | C, ProF, FamF) for different values of the observed variables, but now assuming a dependency between Pro and Fam. Figure 6 indicates that now FamF and (especially) ProF are both relevant in each context. This is different from the case examined in Figure 2 where Pro and Fam were independent and where, once the context was known, considering one dimension was sufficient to infer the identity. In other words, when the independence between Pro and Fam is lost as described in Figure 6, knowing the context does not allow anymore to ignore any of the feature variables.

A dependency among dimensions can also account for conditions where there is a hierarchical relationship between dimensions, being one dimension more abstract than another (an aspect particularly stressed by Self-Categorization Theory; J. C. Turner et al., 1987). An example of this comprises country (with Italy and Spain as categories) and city (with Rome, Florence, Madrid, and Barcelona as categories) as dimensions. Cases like this, where dimensions are hierarchically organized, can be implemented by (i) adding an arrow projecting from the higher-order

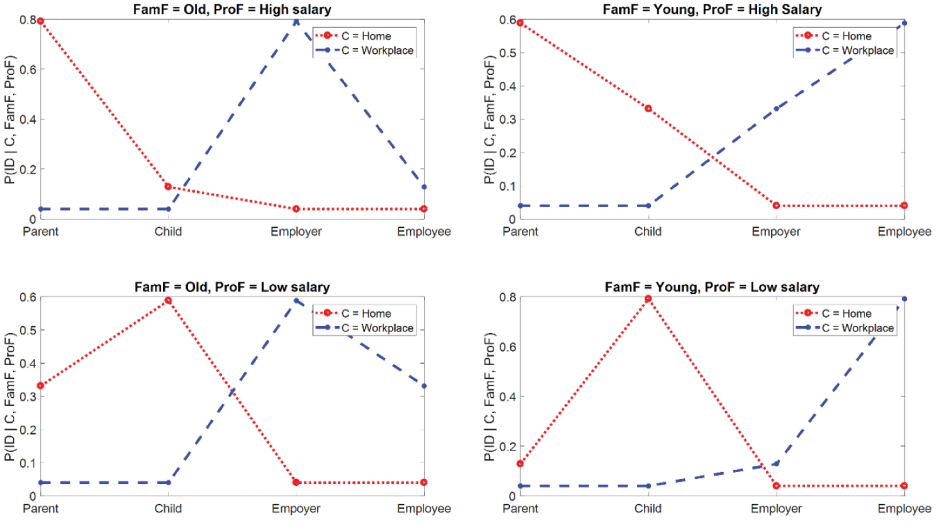

Figure 6. Implications of a model where identity dimensions are not independent (here Fam is influenced by Pro). Posterior of the variable ID (P(ID | C, FamF, ProF)) for different values of C, FamF, and ProF.

dimension (e.g., country) to the lower-order dimension (e.g., city), and by (ii) setting conditional dependencies that never occur (e.g., P(city = Madrid | country = Italy)) to zero. The context variable C can be implemented in such a way that in one condition (e.g., during a world football championship) the higher-order dimension (e.g., country) becomes salient, whereas in another condition (e.g., during a national football championship) the lower-order dimension (e.g., city) becomes salient. This captures the notion, supported empirically (J. C. Turner et al., 1987), that identities at different hierarchical levels can alternate based on the ongoing context.

3.7. Summary

BIMI relies on a simple idea: that identity attribution embodies a form of Bayesian inference. This inference can be based on various observations, including observations of the context and of the target's features. When some of these observations are absent, inference can still be performed. Such cases can be interpreted as forms of prejudice, occurring when an identity is inferred without considering all potentially available evidence. The model offers also an interpretation of stereotypes, occurring when Bayesian inference is employed to guess unobserved target's features. Finally, thanks to the reward variable, BIMI contemplates an influence exerted by motivational factors in identity attribution. Now that BIMI has been explored in some detail, the next section evaluates it in the context of previous theories of identity processing.

4. BIMI and previous theories

This section analyses the implications of BIMI for previous theoretical literature on identity attribution. Broadly speaking, two key features distinguish BIMI from previous proposals. First, previous theories focus on the algorithmic level of analysis. In other words, their approach is inductive: they start by looking at empirical data and then attempt to infer the factors explaining the data. By contrast, BIMI focuses on the functional level of analysis, and thus employs a deductive approach: it starts by formalizing the problem of identity attribution in terms of Bayesian inference, and from this it generates empirical predictions. The second feature characterizing BIMI and absent in previous proposals is the fact that BIMI is expressed in mathematical form. The key advantage of a mathematical formulation is that it reduces ambiguities and enables the identification of precise quantitative predictions.

Besides these general considerations, it is important to compare BIMI against other theories in terms of fitness with empirical evidence. The theoretical literature on identity attribution is vast, and hence the focus

will be restricted to the two prevailing theories in the field: the Social Identity Approach (Hogg, 2020; Tajfel & Turner, 1979; J. C. Turner et al., 1987) and Identity Theory (Burke & Stets, 2009; Stryker & Burke, 2000; Stryker & Serpe, 1982). Before delving in this analysis, it is worth emphasizing that BIMI focuses on a specific aspect of identity, namely on identity attribution (i.e., the mechanisms through which an identity is attributed to the self and others). Research on the concept of identity encompasses also other aspects, such as how identity representations are developed in the first place, and what are the psychological and behavioral consequences of attributing certain identities. These aspects are beyond the scope of BIMI, but they are covered by other theories such as the Social Identity Approach and Identity Theory. Thus, when comparing the BIMI against other theories, the analysis will be confined solely to identity attribution, being this the common ground between BIMI and other theories.

The Social Identity Approach (including Social Identity Theory and Self-Categorization Theory; Hogg, 2020; Tajfel & Turner, 1979; J. C. Turner et al., 1987) is arguably the most well-established tradition for studying identity in social psychology. Its remit is vast, covering a wide range of aspects from how identities develop to how they shape behavior once activated. Regarding the specific issue of identity attribution (which is the focus of BIMI), this has been explored especially by Self-Categorization theory (Hogg, 2020; J. C. Turner et al., 1987). The latter posits three main pillars to explain identity attribution. First, it maintains that identities are highly flexible and context-dependent. Second, it argues that motivation plays an important role in determining which identity is activated. Third, it proposes that available identities are arranged hierarchically, with more abstract and more specific identities occupying higher and lower levels of the hierarchy, respectively, and with the context establishing which hierarchical level is salient. BIMI shares all these three points: specifically, it captures the influence of context thanks to the variable C, the influence of motivation thanks to the reward variable R, and it can implement hierarchical dimensions.

Despite the similarities between BIMI and Self-Categorization theory, the former extends the latter in at least two fundamental ways. First, while Self-Categorization theory implies that only external stimuli determine identity attribution, BIMI also highlight the importance of prior beliefs. Prior beliefs enable BIMI to explore identity attribution in conditions where some features (i.e., about the context or about the target) are not observed. This is proposed to underly the notion of prejudice and stereotype, allowing BIMI to explain these two concepts under an identity attribution framework. A second key difference between Self-Categorization theory and BIMI concerns how different identities interact. In the former approach, at any given moment only one identity is on while all other identities are off (e.g., if

I view another person as an employee, I will discard employer, parent, and child as identities). Conversely, in BIMI each identity is linked with a posterior probability that can be interpreted as reflecting the level of activation of that identity. This allows BIMI to capture nuanced scenarios where multiple identities are, to some extent, evoked simultaneously. For example, if the posterior probability of the employee identity is much higher than the posterior probability of the parent identity, I will treat the other person as an employee, with no ambiguity. However, if the two posterior probabilities are instead close, my attitude will be more ambiguous. This fits with empirical research on identity interference (Oyserman et al., 2006; Settles, 2004), showing that multiple identities not only alternate depending on the context (as stressed by both Self-Categorization theory and BIMI), but can coexist simultaneously (as stressed by BIMI but not by Self-categorization theory).

While the Social Identity Approach is hegemonic for studying identity in social psychology, Identity Theory (Burke & Stets, 2009; Stryker & Burke, 2000; Stryker & Serpe, 1982) is highly influential in microsociology. As for the Social Identity Approach, the scope of Identity Theory is much wider than the issue of identity attribution. Regarding the latter, Identity Theory relies on two key concepts: the idea of identity salience and the idea of identity commitment (Burke & Reitzes, 1991; Burke & Stets, 2009). The theory proposes that, at any given moment, multiple self-identities or self-roles are potentially available, each attached with a salience (reflecting a disposition to activate and enact that identity). This idea is shared by BIMI, where the posterior probability of ID reflects identity salience (in turn indicating a disposition to enact any identity).

The other key concept of Identity Theory relevant for identity attribution is the notion of commitment (Burke & Reitzes, 1991; Burke & Stets, 2009). The theory argues that everyone has different levels of attachment for any identity, depending on factors such as the costs and benefits expected when enacting the identity, and the quality and quantity of social ties associated with it. Identity Theory posits that commitment is an important factor affecting identity salience, with identities imbued with higher commitment being usually also more salient. BIMI captures the idea of commitment thanks to the reward variable R (where the notion of advantage summarizes all aspects underlying commitment): as much as commitment boosts salience in Identity Theory, a more advantageous identity has higher posterior probability in BIMI.

In addition to offering a computational description of salience and commitment, the BIMI clarifies and enrich Identity Theory in many respects. First, while Identity Theory focuses on self-identity, the BIMI encompasses identities attributed to the self and to others alike. Second, the BIMI offers a formal description of the role of context and of the role of

the target's characteristics (e.g., the family and profession dimension) in shaping identity, aspects which remain underexplored in Identity Theory. Third, by considering scenarios where information about context or about target's characteristics is absent, the BIMI offers a formal interpretation of phenomena such as prejudice and stereotype, which are not under the remit of Identity Theory. Finally, while Identity Theory does not fully clarify which factors besides commitment shape salience, the BIMI offers a formal account of how the context and the feature variables partake in this.

In sum, several key aspects highlighted by BIMI fit with and extend previous influential theories of identity in both sociology and psychology. Whether accounts such as Social Identity Theory and Identity Theory are compatible with one another remains debated (Hogg et al., 1995; Stets & Burke, 2000). As discussed above, the BIMI not only appears to capture all essential features of each theory, but it also highlights important aspects disregarded by both. Thus, at least insofar as the issue of identity attribution is concerned, BIMI may offer a platform to reconcile the two previous theories and to extend them further.

5. Conclusions

The paper introduces BIMI, a theory interpreting identity attribution as Bayesian inference. The proposal is that observations about a target's features and about context, combined with motivational factors captured by the reward variable, drive inference about which identity is appropriate in a given situation. An inference can still be made when potentially available information is lacking, a case where prior knowledge about the context and about a target's features becomes critical.

BIMI focuses on simplified scenarios, which allow the model to clarify the basic processes underlying key aspects of identity attribution. However, it is important to stress that this approach comes with important limitations. First, the model focus is restricted to identity attribution, which is only one domain of identity processing. The nature of the (psychological, behavioral, and social) consequences ensuing form attributing a certain identity remains outside the scope of BIMI. Second, BIMI does not address the question of how representations about dimensions (e.g., profession and family), about their categories (e.g., parent versus child), and about prior beliefs develop in the first place. With this regard, a Bayesian proposal (based on structure learning) has been advanced to address this question (Gershman & Cikara, 2020; Lau et al., 2018), positing that dimensions and categories are developed from integrating prior beliefs with experience concerning people's behavior. Third, BIMI presupposes a lack of interaction between context and feature variables, an aspect

which neglects conditions where this interaction is salient. An interesting open question, finally, is how to integrate BIMI within the Active inference and predictive processing framework (Clark, 2013; Friston et al., 2015; Parr et al., 2022; Sprevak, 2021), a proposal that has recently extended Bayesian modeling to account for various complex phenomena underpinning adaptive behavior. An active inference formulation of BIMI may potentially explain more complex scenarios involving a large (potentially infinite) number of possible identities, their relevance and salience, and may offer a more realistic description of how the Bayesian inference processes are embodied in the brain functioning – thus encompassing the implementation level of analysis too.

In conclusion, this paper contributes to the study of identity, with a focus on how identities are attributed. The aim is to explore the issue from a functional angle, focusing on the basic processes involved. Given that Bayesian inference is arguably the most popular functional perspective in psychology, the question asked here is whether any insight can be gained by explaining identity attribution in terms of Bayesian inference. BIMI attempts to address this question. The contribution of the theory is twofold. First, BIMI offers a functional interpretation of identity attribution, highlighting its basic processes in simple mathematical terms. Second, BIMI supports an influential view in psychology advocating a unifying interpretation of mental processes as manifestations of Bayesian inference.

Notes

1. Formally, the joint probability described by the graphical model depicted in Figure 1 can be written as:
   $P(C, ID, R, Fam, Pro, FamF, ProF) = P(C)\ P(Fam)\ P(Pro)\ P(FamF|Fam)\ P(ProF|Pro)\ P(ID|C, Fam, Pro)\ P(R|C,ID)$
2. The code used for simulations is available as supplementary material – this requires Matlab (https://www.mathworks.com) and the Bayesian Network Toolbox (https://www.cs.utah.edu/~tch/notes/matlab/bnt/docs/bnt_pre_sf.html).
3. Formally, the uncertainty is equal to the entropy of the distribution.
   $U = -\sum_i P(ID = i|C, FamF, ProF) log(P(ID = i|C, FamF, ProF))$
4. In social sciences, the notion of prejudice and stereotype is often applied not only when a judgment is made without information (as Allport's definition implies), but also when the judgment is blatantly false, even if the agent examines all information available. Our analysis of BIMI focuses on cases where prejudice or stereotype arise because judgment is made without information (in line with Allport's definition), simply because BIMI offers insight on this. But generally, BIMI is compatible also with cases of prejudice or stereotype based on false judgment: in BIMI, these forms of prejudice and stereotype simply arise when the probability distribution represented in the Bayesian network does not reflect reality.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

No empirical data are analyzed in the paper.

References

Allport, G. W. (1954). The nature of prejudice.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Bissell, K., & Parrott, S. (2013). Prejudice: The role of the media in the development of social bias. *Journalism & Communication Monographs*, *15*(4), 219–270. https://doi.org/10.1177/1522637913504401

Bobo, L., & Kluegel, J. R. (1993). Opposition to race-targeting: Self-interest, stratification ideology, or racial attitudes? *American Sociological Review*, *58*(4), 443–464. https://doi.org/10.2307/2096070

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485–488. https://doi.org/10.1016/j.tics.2012.08.006

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, *17*(5), 475–482. https://doi.org/10.1177/0146167291175001

Burke, P. J., & Reitzes, D. C. (1991). An identity theory approach to commitment. *Social Psychology Quarterly*, *54*(3), 239–251. https://doi.org/10.2307/2786653

Burke, P. J., & Stets, J. E. (2009). *Identity theory*. Oxford University Press.

Burke, P. J., & Tully, J. C. (1977). The measurement of role identity. *Social Forces*, *55*(4), 881–897. https://doi.org/10.2307/2577560

Callero, P. L. (1985). Role-identity salience. *Social Psychology Quarterly*, *48*(3), 203–215. https://doi.org/10.2307/3033681

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540. https://doi.org/10.1006/jesp.1997.1331

Duckitt, J. H. (1992). *The social psychology of prejudice*. Praeger.

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 357–411). McGraw-Hill.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214. https://doi.org/10.1080/17588928.2015.1020053

Gershman, S. J., & Cikara, M. (2020). Social-structure learning. *Current Directions in Psychological Science*, *29*(5), 460–466. https://doi.org/10.1177/0963721420924481

Heise, D. R. (1987). Affect control theory: Concepts and model. *The Journal of Mathematical Sociology*, *13*(1–2), 1–33. https://doi.org/10.1080/0022250X.1987.9990025

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*(1), 189–202. https://doi.org/10.1111/j.1756-8765.2008.01010.x

Hogg, M. A. (2004). Social categorization, depersonalization, and group behavior. In M. B. B. In & M. Hewstone (Eds.), *Self and social identity* (pp. 203–231). Blackwell Publishing.

Hogg, M. A. (2020). *Chapter 5 social identity theory*. Stanford University Press. https://doi.org/10.1515/9781503605626-007

Hogg, M. A., Terry, D. J., & White, K. M. (1995). A tale of two theories: A critical comparison of identity theory with social identity theory. *Social Psychology Quarterly*, *58*(4), 255–269. https://doi.org/10.2307/2787127

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu Rev Psychol*, *55*(1), 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Lau, T., Pouncy, H. T., Gershman, S. J., & Cikara, M. (2018). Discovering social groups via latent structure learning. *Journal of Experimental Psychology: General*, *147*(12), 1881. https://doi.org/10.1037/xge0000470

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Courier Corporation.

Marr, D., & Poggio, T. (1976). From understanding computation to understanding neural circuitry.

McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin*, *87*(1), 195. https://doi.org/10.1037/0033-2909.87.1.195

McGarty, C. (1999). *Categorization in social psychology*. SAGE Publications Ltd. https://doi.org/10.4135/9781446218372

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, *91*(1), 188. https://doi.org/10.1037/0022-3514.91.1.188

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. The MIT Press. https://doi.org/10.7551/mitpress/12441.001.0001

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology*, *25*(1), 57–75. https://doi.org/10.1002/ejsp.2420250106

Reicher, S. (2004). The context of social identity: Domination, resistance, and change. *Political Psychology*, *25*(6), 921–945. https://doi.org/10.1111/j.1467-9221.2004.00403.x

Rigoli, F. (2021). Masters of suspicion: A Bayesian decision model of motivated political reasoning. *Journal for the Theory of Social Behaviour*, *51*(3), 350–370. https://doi.org/10.1111/jtsb.12274

Rigoli, F., Pezzulo, G., Dolan, R., & Friston, K. (2017). A goal-directed Bayesian framework for categorization. *Frontiers in Psychology*, *8*, 408. https://doi.org/10.3389/fpsyg.2017.00408

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

Settles, I. H. (2004). When multiple identities interfere: The role of identity centrality. *Personality and Social Psychology Bulletin*, *30*(4), 487–500. https://doi.org/10.1177/0146167203261885

Sprevak, M. (2021). *Predictive coding I: Introduction*. PhilSci-Archive URL: http://philsci-archive.pitt.edu/id/eprint/19365

Stets, J. E., & Burke, P. J. (2000). Identity theory and social identity theory. *Social Psychology Quarterly*, *63*(3), 224–237. https://doi.org/10.2307/2695870

Stryker, S., & Burke, P. J. (2000). The past, present, and future of an identity theory. *Social Psychology Quarterly*, *63*(4), 284–297. https://doi.org/10.2307/2695840

Stryker, S., & Serpe, R. T. (1982). Commitment, identity salience, and role behavior: Theory and research example. In *Personality, roles, and social behavior* (pp. 199–218). Springer New York. https://doi.org/10.1007/978-1-4613-9469-3_7

Suls, J., & Wheeler, L. Eds. (2013). *Handbook of social comparison: Theory and research*. Springer Science & Business Media. https://doi.org/10.4135/9781446249215.n23

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *Organizational Identity: A Reader*, *56*(65), 9780203505984–16.

Terry, D. J., Hogg, M. A., & White, K. M. (1999). The theory of planned behaviour: Self-identity, social identity and group norms. *British Journal of Social Psychology*, *38*(3), 225–244. https://doi.org/10.1348/014466699164149

Turner, J. C. (1975). Social comparison and social identity: Some prospects for intergroup behaviour. *European Journal of Social Psychology*, *5*(1), 1–34. https://doi.org/10.1002/ejsp.2420050102

Turner, B. M., Dennis, S., & Van Zandt, T. (2013). Likelihood-free Bayesian analysis of memory models. *Psychological Review*, *120*(3), 667. https://doi.org/10.1037/a0032458

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.

Tyler, T. R., & Blader, S. L. (2001). Identity and cooperative behavior in groups. *Group Processes & Intergroup Relations*, *4*(3), 207–226. https://doi.org/10.1177/1368430201004003003

Weigert, A. J., Teitge, J. S., Teitge, J. S., & Teitge, D. W. (2007). *Society and identity: Toward a sociological psychology*. Cambridge University Press.