Volume XVII Annual issue 2024

THE
JOURNAL
OF
PHILOSOPHICAL
ECONOMICS
REFLECTIONS ON ECONOMIC
AND SOCIAL ISSUES

EDITURA
ASE

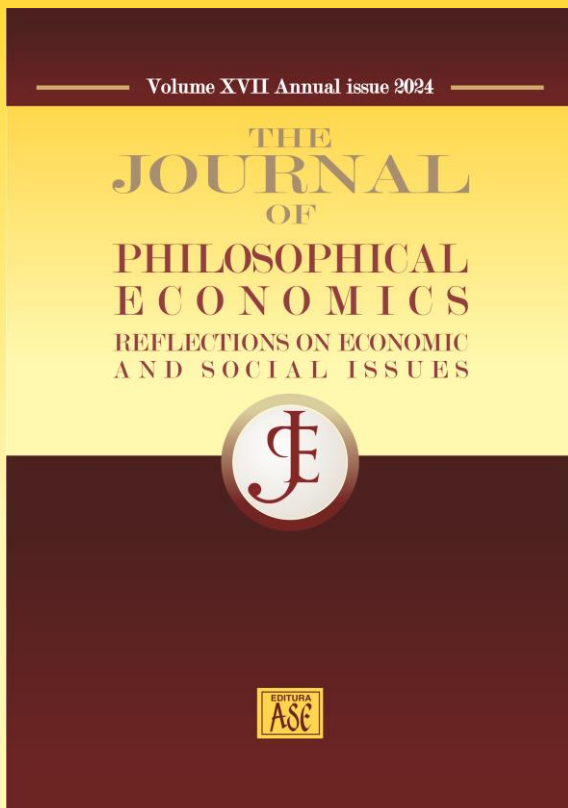# Perspectives on interpersonal utility comparisons: an analysis of selected models

Afschin Gandjour

EDITURA ASE

Perspectives on interpersonal utility comparisons: an analysis of selected models

Afschin Gandjour

Abstract: Recently, new models for comparing the strength of individual preferences have been proposed. This perspective article discusses these models within the context of different accounts of how people attribute mental states to others. The paper highlights that the new models share a common shortcoming with Harsanyi's Equiprobability Model of Moral Value Judgments, which is the inability to facilitate interpersonal comparisons of preference strengths.

Keywords: interpersonal utility comparison; mental states; empathy; sympathy

Introduction

Interpersonal comparison of preferences and preference strengths has been a contentious topic for many decades. Harsanyi (1955, 1977a) introduced a seminal model for interpersonal comparison of strengths of preferences published in 1953, which was extended over the following two decades. The model suggests that Bayesian rationality postulates, together with interpersonal utility comparisons, entail an average utilitarian theory. This model provides an axiomatic foundation for utilitarian morality (Harsanyi 1978), also known as the Equiprobability Model of Moral Value Judgments. It not only considers egoistic preferences but also moral value judgments about the distribution of utility in society. These moral judgments are made by an impartial and rational observer behind a veil of ignorance. The observer has an equal probability of being any member of society. Each member of society possesses von Neumann-Morgenstern (vNM) preferences over lotteries. The observer has preferences over the positions in society that are also represented by a vNM utility function. These preferences are known as extended preferences. They are 'morally valid preferences' and exclude irrational and antisocial preferences (Harsanyi 1975).

Notably, Harsanyi's model is based on empathy as a means of assessing the strength of people's preferences. According to Harsanyi, the observer empathizes with another person, putting themselves in the other person's shoes, and imagining living their life. As a result, the observer completely takes over the preferences of the other person. That is, preferences of the observer and the target become the same, creating a condition Harsanyi refers to as the 'similarity postulate' (1977b, p. 639).

Harsanyi's model has faced several criticisms (see Gandjour 2021, for a recent summary), a central one being that the vNM utility function is not a cardinal representation of utility. In response to this critique, Harsanyi suggested the use of conversion ratios to 'convert all these utility functions into the same common utility unit' (Harsanyi 1977a, p. 57). Thus, the observer is able to perform interpersonal comparisons without the need to map preferences of others onto the observer's own personal scale and using the observer's own personal preferences. The latter implication is desirable as it aligns with what Harsanyi calls the 'principle of acceptance'. He points out that '[t]he interests of each individual must be defined fundamentally in terms of his own personal preferences and not in terms of what somebody else thinks is "good for him"' (Harsanyi 1977a, p. 52). Therefore, the observer must accept the preferences of others.

In a recently published account, Adler (2014) builds upon Harsanyi's Equiprobability Model. Yet, instead of using empathy as a means of conducting intrapersonal or interpersonal utility comparisons, Adler's account relies on sympathy. This approach aims to address two key problems attributed to Harsanyi's account: (i) the satisfaction of certain preferences does not contribute to well-being [1]; and (ii) the empathetic observer may lack information about specific attributes of other individuals, such as birth dates. I will delve into Adler's critique on Harsanyi's account and explore Adler's own model in greater detail.

Another model with the intention of enabling interpersonal comparisons of preferences was presented earlier by Davidson (1986, 2004). According to Davidson, we naturally compare our mental states (beliefs, desires, pretending, knowledge, etc.) with those of other people when interpreting their behaviour. In this process, we project certain aspects of our mental states onto others.

Thus, the basis for interpersonal comparisons is inherent in the very activity of interpretation. However, Weintraub (1998) contends that from interpreting another person's behaviour does not follow that we are able to compare preferences in terms of their strength or intensity. Instead, we are 'force[d] (…) to attribute the same utility scale to all agents; to assume, that is, that agents' utilities straddle the same interval' (p. 309). Hence, according to Weintraub, the model by Davidson does not imply an interpersonal comparison of the intensity of preference satisfaction. Yet, the very same criticism applies to Harsanyi's model using conversion ratios, as discussed earlier, because the aim of conversion ratios is to produce a uniform utility scale for everyone and hence does not account for the intensity of preferences (Weymark 1991).

A relatively recent paper by Rossi (2011) presents a modification of Davidson's model, addressing the criticism raised by Weintraub (1998) and aiming to enable a comparison of individual preference strengths. Specifically, Rossi makes the following case: When we interpret other people's behaviour and attribute mental states to them, we believe that they would form the same preferences and mental states (and, as I infer from his writing, the same strength of preferences) if they were in the same circumstances. Thus, by assuming identical circumstances, we can assess the strength of preferences of others and interpret their behaviour. Still, Rossi argues that this interpretation of behaviour is subject to the *belief* that under the same circumstances people would form the same preferences, which he terms the 'principle of similarity.' If this belief is justified, it becomes possible to make interpersonal comparisons of preference strengths. It is worth noting that the 'principle of similarity' closely resembles Harsanyi's 'similarity postulate' (Harsanyi 1977b).

The aim of this perspective article is to discuss the solutions proposed by Davidson/Rossi and Adler's sympathy-based model within the context of different theories of how people attribute mental states to others. The article highlights that both models are incomplete as they only allow for interpersonal utility comparisons on an ordinal scale, which is also a well-known limitation of Harsanyi's original account. Before delving into the discussion of the two models, the article first presents different theories of mental state attribution.

Accounts of mental state attribution

Two accounts of attributing mental states to oneself and others stand out in the current literature: Theory-Theory accounts (e.g., Carruthers 1996) and Simulation-Theory accounts (e.g., Goldman 2006). Both can be categorized as forms of mindreading, which 'is the activity of representing specific mental states of others, for example, their perceptions, goals, beliefs, expectations, and the like' (Gallese 1998, p. 50). These mental states are 'invoked to explain and predict behaviour' (Gallese 1998, p. 50).

According to Gallese and Goldman (1998, p. 52), the core difference between Theory-Theory and Simulation-Theory accounts lies in their approach to mindreading. Theory-Theory accounts view 'mindreading as a thoroughly "detached" theoretical activity,' where mental states are attributed to others through theoretical reasoning based on causal laws of behaviour. In this process, initial information about the target's beliefs and desires is combined with general principles to generate predictions about their mental states and behaviour. The mental concepts used in this process are drawn from our shared implicit knowledge (Savaki 2010). It is important to note that Theory-Theory accounts do not rely on empathy as a means of attributing mental states to others.

In contrast, simulation theorists argue that we understand others by mentally simulating them. We take the position of the other person, adopt pretend beliefs and pretend desires that we believe the other person has, and use these simulated mental states to comprehend their behaviour (Gallagher 2001, Spaulding 2012). Empathy plays a significant role in simulation theory as it involves empathetically experiencing the mental states of others.

As a word of caution, both accounts of mental state attribution exist in several versions. Theory-Theory accounts vary from strong, rule-based reasoning to more minimal and implicit processes, while Simulation-Theory accounts range from standard simulation to perceptual simulation and interactive simulation. These versions emphasize different aspects of mental state understanding, such as explicit theories, automatic processes, mental simulation, perceptual cues, and the reciprocal interaction between oneself and others (Hutto 2012). In addition, there are hybrid versions that combine the two. Hybrid accounts

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:
Reflections on Economic and Social Issues,* XVII (Annual issue), 1-17

seek to integrate Theory-Theory and Simulation-Theory perspectives in various ways, allowing for a more nuanced understanding of how people understand and make sense of others' thoughts, feelings, intentions, and behaviours. For example, Goldman's simulation theory can be seen as a hybrid Simulation-Theory/Theory-Theory account, as it allows for theory to play a supplementary role (Goldman 2006). On the other hand, some researchers argue for a hybrid Theory-Theory/Simulation-Theory account, with theory playing the dominant role and simulation being supplementary (Nichols and Stich 2003).

The Davidson/Rossi solution

Rossi's account does not address the fundamental processes used by individuals to ascribe mental states to others (and themselves). Instead, it provides a higher-level account of the conditions that enable such attributions. Therefore, based on my interpretation, Rossi's solution can be compatible with different accounts of attributing mental states to others. If Rossi's solution were to invoke empathy as the relevant mechanism, it would result in an account of interpersonal comparisons of preference strengths similar to Harsanyi's (but without the use of conversion ratios). Another way to attribute mental states to others would be through prediction based on implicit knowledge, which is the principle underlying Theory-Theory accounts.

It is important to note that Rossi's 'principle of similarity' begins with 'same circumstances.' However, a comprehensive account of interpersonal utility comparison should also address how utility amounts compare across different circumstances. In other words, we need to compare the utility amount of person i in situation x (as understood by mindreading person i) with the utility amount of person j in situation y (as understood by mindreading person j). However, when comparing preferences across different circumstances, Rossi's account allows for such comparisons only on a ratio scale, not on an absolute scale. [2] A ratio scale establishes the zero point of the utility scale (Bradley 2008), enabling the measurement of relative preference strengths, such as the preference strength for x relative to y (Bradley 2008; see Barrett (2019) for a similar approach based on desire strength). Thus, we can compare utility ratios of different individuals. Bradley (2008) argues that a ratio scale is often

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:*
Reflections on Economic and Social Issues, XVII (Annual issue), 1-17

sufficient, particularly when making aggregate judgments about the relative desirability of two courses of action (i.e., the desirability for x relative to y). [3] According to Bradley (2008), the utility value '0' should be fixed, corresponding to the ethically neutral proposition, which presents a matter of indifference to the individual (Bradley 2008, pp. 95-96). In contrast, the common zero-one rule fixes the utility value '0', corresponding to the worst (Hausman 1995, p. 480) or least preferred option. Bradley's proposal differs from the common zero-one rule because the ethically neutral proposition does not align with an individual's least preferred option (Bradley 2008, p. 96). Bradley explicitly allows for prospects that are less desirable than the ethically neutral proposition by assigning them negative utility values.

But regardless of how the zero point is operationalized, from a utilitarian perspective, which aims to assess the aggregated utility of a specific course of action, it is still crucial to represent the strength of preferences on an absolute scale. This absolute scale is unique to each individual (cf. Davidson 1955). The reason for this is that a utilitarian seeks to understand to what extent the satisfaction of one person's preference can compensate for the dissatisfaction of another person's preference. When conducting this analysis in comparison to the status quo, it is equivalent to comparing the utility gains and losses (utility increments) relative to the status quo (Narens and Skyrms 2020).

Rossi's solution differs from Harsanyi's in another respect. According to Harsanyi, the 'similarity postulate' is justified based on pragmatic considerations. It is considered the simplest, most parsimonious, and least arbitrary hypothesis among alternative explanations (Harsanyi 1982, p. 51). On the other hand, Rossi's 'principle of similarity' is the principle that enables the interpretation of other people's behaviour. If that is the case, the arguments put forth by Rossi and Harsanyi to defend the possibility of justified interpersonal utility comparisons are different: Rossi presents a 'modest' transcendental argument, while Harsanyi offers a pragmatic argument.

Sympathy-based account

In line with other authors (Parfit 1984, p. 494; Scanlon 1996; Arneson 1999, p. 124; Darwall 2002, p. 53), Adler (2014) highlights that satisfying preferences

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:
Reflections on Economic and Social Issues,* XVII (Annual issue), 1-17

does not always contribute to well-being. Observers can have 'non-self-interested preferences.' Adler refers to this as the 'wrong kind of preference' problem. He provides the following example, which is cited in full because it is an important (but mistaken) reason for his alternative proposal discussed below:

> [I]magine that there are five people in the population: *i, j, k, l, m*. Outcome *x* is one in which individuals' incomes range in $20,000 increments from $20,000 to $100,000. Individual *i* has income $20,000, and individual *j* has income $100,000; while the other three have, respectively, incomes of $40,000, $60,000 and $80,000. Individual *i* has tastes R_i, etc.

> Then $(A_i(x),\ R_i)$ is the bundle (having an income of $20,000; having tastes R_i; being part of a population of five individuals where the other incomes are $40,000, $60,000, $80,000, $100,000 and where the other individuals have tastes R_k, R_l, R_m and R_j). And $(A_j(x),\ R_j)$ is the bundle (having an income of $100,000; having tastes R_j; being part of a population of five individuals where the other incomes are $20,000, $40,000, $60,000, $80,000 and where the other tastes are R_i, R_k, R_l and R_m).

> Imagine, now, that *k* is an *impartial* spectator. In the exercise of ranking hybrid bundles, she assumes an attitude, not of self-interest, but rather of impartiality between her interests and everyone else's. If so, *k* will be *indifferent* between the bundles $(A_i(x),\ R_i)$ and $(A_j(x),\ R_j)$. She doesn't care, from this impartial perspective, whether she is the one with $20 000 and particular tastes in a given population distribution of income and tastes, or she is the one with $100,000 and particular tastes in the very same distribution of income and tastes. But, of course, (*x; i*) and (*x; j*) are not equally good for well-being. It is worse for well-being, ceteris paribus, to be the person with the lowest income in a given distribution of income, rather than the person with the highest (at least if R_i and R_j both include a taste for more income rather than less).

The other significant problem identified by Adler (2014) in Harsanyi's account is that individuals may possess attributes that the observer inherently lacks and cannot acquire without fundamentally changing who they are. Therefore, the observer cannot truly put themselves in the individual's shoes. Adler presents the example of an observer who is tasked with forming extended preferences over two lives, one in the first century BC and the other in the 16th

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:
Reflections on Economic and Social Issues,* XVII (Annual issue), 1-17

century AD. He argues that an observer born in 1980 inherently lacks the essential attribute of birth timing in those two lives.

Due to these reasons, Adler (2014) proposes a 'sympathy-based conception of extended preferences.' In this conception, when ranking an individual's outcomes [4] x and y, the observer 'does *not* engage in the thought experiment of acquiring' an individual's causal factors or 'attributes' associated with x and y. Instead, the observer's extended preference is reduced to an outcome preference under a condition of unreserved sympathy [5] for the individual (while still being represented by a vNM utility function). By reducing the preference exercise to an assessment of outcomes, the observer does not need to consider non-mental attributes such as birth dates. However, the observer 'can take account of all of the subjects' attributes (…) in arriving at [his] well-being judgements.' The observer can take account of the individual's preferences, 'without requiring the observer to take those preferences as decisive.' From my interpretation, the observer may choose to disregard certain attributes, causal factors, and background information of an individual, even if they are known. In such cases, the observer may suppress some background information of the individual and perhaps incorporate their own background information. Nonetheless, the observer is never required to imagine acquiring someone else's identity (p. 158).

Adler's perspective that sympathy does not necessitate mindreading through empathy finds support in the literature. For instance, Sober and Wilson (1998, p. 236) have argued that empathy requires one to be a psychologist, whereas sympathy does not:

> Empathy entails a belief about the emotions experienced by another person. Empathic individuals are 'psychologists' (..); they have beliefs about the mental states of others. Sympathy does not require this. You can sympathize with someone just by being moved by their objective situation; you need not consider their subjective state. Sympathetic individuals have minds, of course; but it is not part of our definition that sympathetic individuals must be psychologists.

Similarly, Stueber (2018) argues that 'sympathy does not necessarily require feeling any kind of congruent emotions on part of the observer, a detached recognition or representation that the other is in need or suffers might be

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:*
Reflections on Economic and Social Issues, XVII (Annual issue), 1-17

sufficient.' Based on my interpretation, Adler's approach could therefore be classified as a Theory-Theory account.

Furthermore, Adler distinguishes between paternalistic [6] and non-paternalistic sympathetic preferences. In fact, his account can accommodate both types of preferences. While paternalistic altruism is utility derived from another's consumption, non-paternalistic altruism is utility derived from another's own utility. As vividly described by Hoffmann (2006):

> Parents have paternalistic concern for their children when they care about their children's health or consumption in and of itself, not because of what the child likes. A classic example of paternalistic caring is the parent's admonishment, 'Eat your spinach. I don't care if you don't like it. It's good for you.' Parents have non-paternalistic concern for their children when they care about the child's consumption or health because it makes the child happy.

Adler argues that limiting his account to non-paternalistic preferences would result in the same ranking of outcomes by the observer and the subject. [7] In support of his reasoning, economists now widely accept that non-paternalistic altruism leads to double counting of individual utility and should hence be excluded from preference elicitation exercises (Bergstrom 1982). This is because an individual utility is already counted once in the utility function of the individual in question and counting it again in the utility function of the individual demonstrating non-paternalistic altruism would result in double counting (Bergstrom 1982). The argument is based on the observation that, with non-paternalistic altruism, the necessary conditions for optimality remain the same as when only private valuation is considered (Bergstrom 1982). Therefore, compared to self-interested preferences, non-paternalistic altruism does not alter the allocation of resources or the ranking of outcomes.

In the following I provide a critique of Adler's 'sympathy-based conception.' First, his motivating example cited above is mistaken. It does not represent an example of the 'wrong kind of preference' problem. The literature cited by Adler in reference to the 'wrong kind of preference' problem (Parfit 1984, p. 494; Scanlon 1996, Arneson 1999, p. 124; Darwall 2002, p. 53) exclusively concerns *egoistic* preferences that do not contribute to well-being. However, in his example, Adler criticizes the fact that satisfying *moral* preferences does not

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:
Reflections on Economic and Social Issues,* XVII (Annual issue), 1-17

contribute to well-being. Contrary to Adler, Harsanyi (1982) argues that this is actually desirable. He explains (Harsanyi 1982): 'Otherwise [the observer's] assessment will not be a genuine moral value judgement but rather will be merely a judgement of personal preference.' In Harsanyi's account, satisfying moral preferences does not contribute to the well-being of the observer. Instead, moral preferences are intrinsically important. Therefore, it is not plausible to criticize Harsanyi's account for including moral preferences that do not directly contribute to well-being.

My second point of criticism concerns the use of paternalistic preferences in Adler's account. In Harsanyi's account, preferences of individuals are 'excluded' if they are irrational (Harsanyi 1982). Chang (2000) considers this to be a 'paternalistic intervention to promote a person's own good.' Furthermore, the exclusion of irrational preferences in Harsanyi's account needs to be justified from a consequentialist viewpoint (Birnbacher, unpublished lecture notes). Merely arguing that irrational preferences have inherent low value is an insufficient reason and a criticism of Harsanyi's account (Birnbacher, unpublished lecture notes). In contrast to Chang's (2000) interpretation of Harsanyi's account, Adler does not discuss paternalistic preferences in relation to rational preferences. Nor does he discuss paternalistic preferences in relation to moral preferences. Therefore, based on Adler's own presentation, the second point of criticism is unrelated to the first (i.e., the non-use of moral preferences in Adler's account). Specifically, I argue that Adler's account needs to present an underlying coherent theory in the first place that justifies overriding the preferences of individuals. Adler could have made reference to so-called moral paternalism, which aims to promote the moral well-being of a person (Dworkin 2020). However, given Adler's criticism of the inclusion of moral preferences, this would have led to a contradiction. Alternatively, Adler could have turned to welfare paternalism, which accommodates a regard for the welfare of another individual. However, this does not align with the rational preferences by the observer, which do not require additional correction. Therefore, Adler's account faces a dilemma: it is either based solely on non-paternalistic preferences and yields the same result as an empathy-based assessment of outcomes that *matter to the subject*, or it includes paternalistic preferences without an independent theoretical foundation.

Merely stating that, from behind the veil of ignorance, non-paternalistic preferences yield the same result as an empathy-based assessment of outcomes cannot, by itself, justify to reverting to paternalistic preferences. Furthermore, reducing the preference exercise to a pure outcome assessment appears to be an ad hoc fix aimed at precluding the problems associated with Harsanyi's account. Instead, there needs to be a coherent theory providing a rational justification for using sympathy and its underlying psychological mechanism as a means of conducting interpersonal comparisons. The theoretical justification for using sympathy and outcomes would need to be embedded in Harsanyi's axiomatic justification of utilitarianism.

In addition, a sympathy-based account that simply ignores all or only missing attributes runs the risk of introducing bias. This is because the assessment of an individual's outcomes can depend on the underlying attributes. Our unreserved sympathy for another person's outcomes relies on the available background information about that person. For instance, the assessment of an individual's health status is likely influenced by the time period in which they lived. A functional limitation may be perceived as less concerning when evaluating the profile of a person from earlier times compared to today, considering the advancements in health technologies and resource availability. Thus, the absence of information on attributes such as birth timing could lead to biased assessments. Hédoin (2021) does not consider this to be a relevant problem because Adler does not seek 'uniformity of extended preferences.' However, this argument implicitly assumes that biases cancel out across different observers.

A final point of critique on Adler's sympathy-based approach is that it can face criticism on similar grounds as Harsanyi's account regarding the use of the vNM utility function. This criticism applies whether Adler's approach is used for interpersonal comparisons of welfare for their own sake or for utility maximization. According to Weymark (1991), the vNM utility function used in Harsanyi's original model allows only for an ordinal ranking of preferences, but not for an interpersonal comparison of utility that satisfies 'cardinal unit plus comparability' (a point also stated by Sen 1986). However, only the latter scale can capture differences in preference intensity. In fact, Broome (2008) considers this to be the 'standard objection' to Harsanyi's model. [8] It implies

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:*
Reflections on Economic and Social Issues, XVII (Annual issue), 1-17

that the social welfare function is linear only in terms of individual von Neumann-Morgenstern utilities, but not in terms of welfare (Weymark 1991, p. 313). Therefore, Harsanyi's theorem merely calls for maximizing the sum of vNM utilities of individuals, but it should not be interpreted as a utilitarian theorem or as supporting utilitarianism. Harsanyi's suggestion to use conversion ratios to establish an interpersonal comparison of preference strengths through a 'common utility unit' (Harsanyi 1977a, p. 56) also fails because this information is not deducible from the vNM utility functions of individuals. Similarly, a ratio scale, which would establish cardinality but is not invoked by Harsanyi or Adler, is not implied by the vNM utility function, and would require a different conceptual foundation.

Conclusions

This perspective article aims to explore the Davidson/Rossi solution and Adler's sympathy-based model in the context of various accounts of mental state attribution. The Davidson/Rossi solution focuses on the conditions that allow for attributing mental states but does not address the underlying processes. Rossi's principle of similarity enables interpersonal utility comparison, but it only allows for comparison on a ratio scale, not an absolute scale. The principle can be compatible with both Theory-Theory and Simulation-Theory accounts, depending on whether empathy or implicit knowledge is used in mental state attribution.

Adler's account is sympathy-based, which suggests that extended preferences can be ranked by assessing outcomes under conditions of unreserved sympathy. This account avoids considering missing or irrelevant attributes of individuals and focuses solely on outcomes. However, this approach raises concerns about potential biases and lacks a coherent theoretical foundation. Additionally, both Adler's sympathy-based account and Harsanyi's empathy-based account suffer from criticisms related to the von Neumann-Morgenstern utility function, as they fail to provide a basis for cardinal interpersonal utility comparison.

Endnotes

[1] According to Adler (2014), '[w]ell-being, conceptually, involves subject-relative goodness.'

[2] A cardinal or metric scale allows for the measurement of differences and/or proportions of the outcomes of the characteristic of interest. There are three different types of cardinal scales: interval, proportional (ratio), and absolute scale (Mittag and Horst 1993). Both the proportional (ratio) and the absolute scale start at the natural origin zero (Mittag and Horst 1993). In contrast to a proportional (ratio) scale, however, an absolute scale has natural units (Mittag and Horst 1993).

[3] From my interpretation of Bradley's account, relative desirability is not an ordinal measure because relative desirability is quantifiable.

[4] According to Adler, outcomes are defined as 'arbitrarily detailed specifications of possible worlds, [which] do not specify individuals' preferences.'

[5] According to the Stanford Encyclopedia of Philosophy, sympathy is defined as follows (Stueber 2018): 'In contrast to affective empathy, sympathy is not an emotion that is congruent with the other's emotion or situation such as feeling the sadness of the other person's grieving for the death of his father. Rather, sympathy is seen as an emotion sui generis that has the other's negative emotion or situation as its object from the perspective of somebody who cares for the other person's wellbeing (Darwall 1998). In this sense, sympathy consists of 'feeling sorrow or concern for the distressed or needy other,' a feeling for the other out of a 'heightened awareness of the suffering of another person as something that needs to be alleviated'.

[6] Making decisions for other people rather than letting them take responsibility for their own lives (Cambridge Dictionary).

[7] 'Strong non-paternalism says that the well-being ranking of a given subject's histories is identical to the subject's extended preferences over those histories.' (Adler 2014, p. 155)

[8] In contrast, Hausman (1995) argues that it is unnecessary, or even incorrect, for preference utilitarianism to establish an absolute unit of satisfaction. He contends that an absolute scale would have to account for the effect of changes in the level of preference satisfaction on mental states. But if preference utilitarianism is a theory focused on preferences, then the impact on mental states is morally irrelevant.

Conflict of Interest Statement

The author declares that there is no conflict of interest.

References

Adler, M. D. (2014), Extended preferences and interpersonal comparisons: a new account', *Economics and Philosophy*, 30(2), 123-162.

Arneson, Richard J. (1999), 'Human flourishing versus desire satisfaction', in Ellen F. Paul, Fred D. Miller, and Jeffrey Paul (eds.), *Human Flourishing*, Cambridge: Cambridge University Press, pp. 113-142.

Barrett, J. (2019), 'Interpersonal comparisons with preferences and desires', *Politics, Philosophy & Economics*, 18(3), 219-241.

Bergstrom, Theodore C. (1982), 'When is a man's life worth more than his human capital?', in Michael W. Jones-Lee (ed.), *The Value of Life and Safety*, Amsterdam: North-Holland, pp. 3-26.

Bradley, R. (2008), 'Comparing evaluations', *Proceedings of the Aristotelian Society*, 108, 85-100.

Broome, John (2008), 'Can there be a preference-based Utilitarianism?', in Marc Fleurbaey, Maurice Salles, and John A. Weymark (eds.), *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, Cambridge: Cambridge University Press, pp. 195-211.

Cambridge Dictionary. 'Paternalistic'. Online at https://dictionary.cambridge.org/de/worterbuch/englisch/paternalistic.

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:
Reflections on Economic and Social Issues,* XVII (Annual issue), 1-17

Carruthers, Peter (1996), 'Simulation and self-knowledge: a defense of the Theory-Theory,' in Peter Carruthers and Peter K. Smith (eds.), *Theories of Theories of Mind*, Cambridge: Cambridge University Press, pp. 140-160.

Chang, H-J. (2000), 'A liberal theory of social welfare: fairness, utility, and the Pareto principle', *The Yale Law Journal*, 110(2), 173-235.

Darwall, Stephen L. (2002), *Welfare and Rational Care*, Princeton, NJ: Princeton University Press.

Davidson, D., McKinsey, J. C., and Suppes, P. (1955), 'Outlines of a formal theory of value, I', *Philosophy of Science*, 22(2), 140-160.

Davidson, Donald (1986), 'Judging interpersonal interests', in Jon Elster and Aanund Hylland (eds), *Foundations of Social Choice Theory*, Cambridge: Cambridge University Press, pp. 195-211.

Davidson, Donald (2004), *Problem of Rationality*, Oxford: Clarendon Press.

Dworkin, Gerald (2020), 'Paternalism', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition). Online: https://plato.stanford.edu/archives/fall2020/entries/paternalism/.

Gallagher, S. (2001), 'The practice of mind: theory, simulation or primary interaction?', *Journal of Consciousness Studies*, 8(5-6), 83-108.

Gallese, V., and Goldman, A. I. (1998), 'Mirror neurons and the simulation theory of mind-reading', *Trends in Cognitive Sciences*, 2(12), 493-501.

Gandjour, A. (2021), 'Interpersonal comparison of welfare based on Harsanyi's equiprobability model for moral value judgments', *Ethical Perspectives*, 28(4), 385-416.

Goldman, Alvin I. (2006), *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*, New York: Oxford University Press.

Harsanyi, J. C. (1953), 'Cardinal utility in welfare economics and in the theory of risk-bearing', *Journal of Political Economy*, 61, 434-435.

Harsanyi, J. C. (1955), 'Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility', *Journal of Political Economy*, 63, 309-321.

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:
Reflections on Economic and Social Issues,* XVII (Annual issue), 1-17

Harsanyi, J. C. (1975), 'Preferences and utilitarian theory: some comments', *Erkenntnis*, 13(3), 397-399.

Harsanyi, John C. (1977a), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

Harsanyi, J. C. (1977b), 'Morality and the theory of rational behavior', *Social Research*, 44, 623-656.

Harsanyi, J. C (1978), 'Bayesian decision theory and utilitarian ethics', *American Economic Review*, 68(2), 223-228.

Harsanyi, John C. (1982), 'Morality and the theory of rational behaviour', in Amartya Sen and Bernard Williams (eds.), *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, pp. 39-62.

Hausman, D. M. (1995), 'The impossibility of interpersonal utility comparisons', *Mind*, 104(415), 473-490.

Hédoin, C. (2021), 'Social contract, extended goodness, and moral disagreement', *Erasmus Journal for Philosophy and Economics*, 14(2), 25-52.

Hoffmann, S. (2006), 'Since children are not little adults - socially - what's an environmental economist to do', *Duke Environmental Law & Policy Forum*, 17, 209-232.

Hutto, Daniel D. (2012), *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*, MIT Press.

Mittag, Hans J., Rinne, Horst (1993), *Statistical Methods of Quality Assurance*, London: Chapman & Hall.

Narens, L., Skyrms, B. (2020), *The Pursuit of Happiness: Philosophical and Psychological Foundations of Utility*, Oxford: Oxford University Press.

Nichols, S., Stich, S. P. (2003), *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford University Press.

Parfit, Derek (1984), *Reasons and Persons*, Oxford: Clarendon Press.

Gandjour Afschin (2024), Perspectives on interpersonal utility comparisons:
an analysis of selected models, *The Journal of Philosophical Economics:*
Reflections on Economic and Social Issues, XVII (Annual issue), 1-17

Rossi M. (2011), 'Transcendental arguments and interpersonal utility comparisons', *Economics and Philosophy*, 27, 273-295.

Savaki, H. E. (2010), 'How do we understand the actions of others? By mental simulation, not mirroring', *Cognitive Critique*, 2, 99-140.

Scanlon T. M. (1996), 'The status of well-being'. Online: http://tannerlectures.utah.edu/_documents/a-to-z/s/Scanlon98.pdf (accessed March 19, 2020).

Sen, Amartya K. (1986), 'Social choice theory', in Arrow, Kenneth, Intriligator, Michael (eds.), *Handbook of Mathematical Economics*, Volume 3, North Holland, pp. 1073-1181.

Sober, Elliott, Wilson, David S. (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Harvard University Press.

Spaulding, S. (2012), 'Mirror neurons are not evidence for the simulation theory', *Synthese*, 189, 515-534.

Stueber K. (2018), 'Empathy', in Zalta E. N. (ed.), *The Stanford Encyclopaedia of Philosophy* (Spring 2018 Edition). Online: https://plato.stanford.edu/archives/spr2018/entries/empathy/ (accessed March 19, 2020).

Weintraub, R. (1998), 'Do utility comparisons pose a problem?', *Philosophical Studies*, 92, 307-319.

Weymark, John A. (1991), *'A reconsideration of the Harsanyi-Sen debate on utilitarianism'*, in Elster, Jon, Roemer, John E. (eds.), *Interpersonal Comparisons of Well-being*, Cambridge: Cambridge University Press, pp. 255-320.

Afschin Gandjour is Professor with the Frankfurt School of Finance & Management, Frankfurt am Main (Germany) (a.gandjour@fs.de).