

# Aussagekraft sensorischer Urteile

**Die Aussagekraft sensorischer Ergebnisse hängt von der Qualifikation der Koster ab. Durch statistische Verfahren kann sie abgesichert werden. Qualität ist subjektiv und entzieht sich einer Normung.**

Die sensorische Bewertung lässt sich unter vielen Aspekten mit den Verfahren der chemischen oder physikalischen Analyse von Stoffgemischen vergleichen. Sie ist ein Messverfahren, bei dem der Mensch als Messinstrument eingesetzt wird. Es resultiert ein Messergebnis, welches in Form von Zahlen ausgedrückt wird. Solche Zahlen können die Intensitätswerte definierter sensorischer Merkmale im Rahmen technischer Auswertungen darstellen. Geläufiger sind sie als Punkte, mit denen ein Wein bei der Qualitätsweinprüfung oder bei Wettbewerben bewertet wird. Zuverlässigkeit und Aussagekraft derartiger Messergebnisse hängen zwangsläufig von der Qualität der als Sensoren eingesetzten Prüfer ab. Insofern erfordert die Bewertung von Wein auch eine Bewertung der Prüfer.

## Qualität der Prüfer

Leider funktioniert der Mensch als Messinstrument weniger zuverlässig als ein chemischer Detektor im Labor, weil psychologische Momente und physiologische Umstände seine sensorische Leistungsfähigkeit beeinflussen. Allein aufgrund ihrer genetischen Variabilität sprechen Menschen unterschiedlich stark auf bestimmte Stimuli an. Selbst ge-

durch Schulung verbessert werden.

In der chemo-physikalischen Messtechnik werden Methoden unter den Aspekten von Spezifität, Sensibilität

*Es ist eines der Grundübel der Weinbranche, dass oft Erzeuger unter sich ihre Produkte bewerten und sich gegenseitig beweihräuchern.*

und Reproduzierbarkeit beurteilt. Ähnliche Anforderungen werden in der Sensorik an die Qualifikation der Prüfer gestellt. Sie wird nach folgenden drei Kriterien bemessen:

### Diskriminiervermögen

Darunter versteht man die Befähigung, spezifische Sinneseindrücke eindeutig zu erkennen, selektiv zu benennen und zwischen ihnen unterscheiden, diskriminieren, zu können. Auf Wein angewandt heißt das, dass ein Prüfer zwischen UTA und Böckser, zwischen bitter und sauer, zwischen fruchtigen und vegetabilen Aromen unterscheiden können sollte. Das Diskriminiervermögen eines Prüfers ist auch an die Beherrschung eines präzisen Vokabulars zur Benennung der Sinneseindrücke gebunden.

In der Regel ist das sensorische Diskriminiervermögen besser als das dazugehörige verbale Ausdrucksvermögen.

Der Prüfer erkennt Eigenschaften von Weinen und Unterschiede zwischen Weinen, kann ihnen aber nicht immer die entsprechenden Begriffe zuordnen. Daraus resultiert das

sprachliche Problem in der Sensorik. Da das Diskriminiervermögen an die Beherrschung sprachlicher Instrumente geknüpft ist, kann es durch Schulung nur teilweise verbessert werden.

### Sensibilität

Sie beschreibt die Fähigkeit, gegebene Stimuli in möglichst niedriger Konzentration wahrzunehmen und geringste Konzentrationsunterschiede zu erfassen. Selbst zwischen geschulten Prüfern können die individuellen Unterschiede in der Geruchsemp-

findlichkeit um über eine Zehnerpotenz streuen. Sie nimmt im Alter ab und legt eine periodische Überprüfung nahe. Im Weinbereich wird die Sensibilität der Prüfer meist durch Bemessung ihrer Geschmacks- und Differenzschwellenwerte für süß, sauer und bitter anhand wässriger Lösungen von Saccharose, Zitronensäure und Chininchlorid ermittelt. Leider sind komplexe Geruchseindrücke oder gar Weinfehler schwierig zu standardisieren. Sensibilität ist überwiegend eine Frage von Training und kann durch Schulung verbessert werden.

### Reproduzierbarkeit

Unter Reproduzierbarkeit versteht man die Fähigkeit, bei wiederholter Verkostung eines Weins oder einer Testlösung unabhängig von Zeitpunkt, Probenfolge und Tagesform zu einem möglichst gleichen Urteil zu gelangen. Sie lässt sich überprüfen, indem man bei Verkostungen Kontrollproben mitlaufen lässt oder unbekannte, aber identische Proben mehrfach vorstellt. Je größer die Abweichung zwischen den Einzelbeurteilungen der gleichen Probe, desto geringer ist die Reproduzierbarkeit des Prüfers. Unsichere Verkoster, die in Wiederholung vorgestellte identi-

*Man kann ohne Übertreibung sagen, dass die Sensorik mit der Statistik lebt oder stirbt.*

schulte und ausgewählte Prüfer können Fehlurteile zustande bringen. Deshalb wurden Verfahren entwickelt, mit denen Qualität und Funktionsweise der Prüfer gemessen oder

sche Weine stark unterschiedlich bewerten, sind ein seriöses Problem in der Sensorik. Reproduzierbarkeit ist überwiegend eine Frage von Erfahrung. Durch sie lernt der Prüfer, die auf ihn einwirkenden psychophysikalischen Einflüsse zu erkennen und zu kontrollieren. Erfahrene Prüfer können sogar recht gut ihre Tagesform einschätzen.

**Schwellenwerttests und Wiederholungen** sind die gängigsten Verfahren zur Beurteilung der Prüferqualifikation. Mittels standardisierbarer Referenzlösungen lassen sich Prüfer auf definierte Stimuli eichen. Es ist sogar möglich, für jeden Koster ein Sensibilitätsdiagramm aufzustellen. Dieses Vorgehen entspricht dem völligen Durchmessen eines Prüfers. Da sich die Schwellenwerte mit der Zeit verändern, muss ein Sensibilitätsdiagramm periodisch überprüft und gegebenenfalls korrigiert werden. Der damit verbundene Zeitaufwand gibt diesem Verfahren einen fast futuristischen Anstrich, obgleich er im Rahmen sensorischer Forschungsaufgaben unvermeidbar werden kann.

### Beurteilung von Prüfungsergebnissen

Jeder Prüfer hat seine persönlichen Stärken, Schwächen und Toleranzgrenzen. Sensorische Universalisten gibt es nicht. Gute Sensoriker wissen ihr eigenes Prüfverhalten einzuschätzen.

*Jeder Prüfer hat seine persönlichen Stärken, Schwächen und Toleranzgrenzen. Sensorische Universalisten gibt es nicht.*

zen. Die besten Prüfer sind erfahrungsgemäß die, welche sich ihrer eigenen Schwächen bewusst sind und sich immer wieder selbst in Frage stellen. **In der Sensorik ist Bescheidenheit eine Tugend.**

Da Koster höchst variabel auf einzelne Stimuli reagieren und Qualität unterschiedlich wahrnehmen, wird zum Zweck sensorischer Auswertungen üblicherweise eine Gruppe von Prüfern eingesetzt. Damit stellt sich die berechnete Frage nach der Aussagekraft der sich ergebenden Urteile. Ähnlich wie sich die Qualifikation einzelner Koster überprüfen lässt, kann man mit einigen mathematischen Grundlagen auch die Zuverlässigkeit sensorischer Prüfungsergebnisse beurteilen.

Üblicherweise wird der negative Einfluss abweichender Einzelprüfer auf das Prüfungsergebnis durch die Errechnung von Durchschnittswerten umgangen. Dabei treten Unterschiede zwischen den Proben auf, die real sind oder nur durch die Streuung zwischen den Prüfern vorgetäuscht werden. Der Mittelwert als solches gibt keine Information über die Präzision des Ergebnisses. Eine erste Einschätzung seiner Aussagekraft kann man durch die Standardabweichung erhalten. Darunter versteht man die durchschnittliche Streuung der Einzelergebnisse um den Mittelwert. Die systematische Ausweisung der Standardabweichung, bei chemisch-physikalischen Messungen längst üblich, würde auch eine wichtige Information über den Wert sensorischer Urteile geben.

In der Praxis kommt es immer wieder vor, dass Sensoriker identische, in Wiederholung vorgestellte Weine stark unterschiedlich bewerten. Solche durch unsichere Verkoster bedingte Bewertungsdifferenzen ergeben sich aus der Tagesform. Sie erschweren die Aussage und sind zur Bildung eines Mittelwertes nicht geeignet. Ebenso findet man Einzelprüfer, die zwar gegenläufig zum allgemeinen Trend urteilen, aber ihre Aussage zuverlässig wiederholen. Solche Prüfer sind aufgrund ihrer Reproduzierbarkeit sehr wertvoll, vergrößern aber zwangsläufig die Standardabweichung. In beiden Fällen können Unterschiede zwischen Weinen vorgetäuscht oder verwischt werden.

**Varianzanalyse sichert Ergebnisse ab**

Die Varianzanalyse ist das mathematische Instrument, mit dem reale Unterschiede von scheinbaren Unterschieden abgegrenzt werden. Sie überprüft, mit welcher statistischen Sicherheit eine Aussage getroffen werden kann und welche Faktoren in welchem Maß zu der Schwankungsbreite (Varianz) von sensorischen Ergebnissen beitragen. Entsprechende Rechenprogramme laufen auf allen PC's.

Im günstigsten Fall trägt allein der Unterschied zwischen den Weinen in signifikantem Maß zur Varianz bei (Faktor „Wein“). Das setzt voraus, dass ihn alle beteiligten Prüfer gleich wahrnehmen. Diese unterscheiden sich aber oftmals in der Benutzung der Bewertungsskala, indem einige die volle Skalenlänge ausnutzen, während andere vorsichtiger bewerten und sich vor dem Gebrauch höherer Noten scheuen. In der Praxis fließt das Prüferverhalten mit in die Varianz ein (**Faktor „Prüfer“**). Kommt es zusätzlich zu einer Wiederholung der Probe, wird die Bewertung eines jeden Prüfers mehr oder weniger stark von seiner Erstbewertung abweichen. Das Urteil ist nicht nur vom Wein, sondern auch vom Zeitpunkt abhängig (**Faktor „Zeit“**).

Die durch die Faktoren „Prüfer“ und „Zeit“ hervorgerufenen Streuungen addieren sich. Beurteilen zwei Prüfer einen Wein zum gleichen Zeitpunkt und geben die Intensität ihrer Wahrnehmung anhand einer Punkteskala von 0 bis 10 wider, so mag Prüfer A flüchtige Säure mit einer Intensität von 9 und Prüfer B mit einer Intensität von nur 6 wahrnehmen. Der Unterschied zwischen den Prüfern entspricht drei Intensitätspunkten. Bewertet nun Prüfer B den gleichen Wein zu einem anderen Zeitpunkt, mag er eine Intensität von nur 4,5 angeben. In diesem Fall addiert sich der durch den Faktor Zeit zu erklärende Unterschied (1,5) zu dem durch die Prüfer bedingten Unterschied (3) zu einer absoluten Differenz von 4,5 Punkten. Prüfer A hat die flüchtige Säure doppelt so stark wahrgenommen als Prüfer B und wird den Wein möglicherweise ablehnen, während Prüfer B ihn noch akzeptiert.

Dieses kumulative Verhalten von Streuungen ist verantwortlich für anscheinend widersprüchliche Beurteilungen eines Weines durch unterschiedliche Prüfungskommissionen, wenn dabei die Grenze zwischen

*Obwohl Einzelbewertungen (Weinpäpste) von zweifelhaftem Wert sind, wird Qualität durch die Autorität einer einzelnen Person legitimiert.*

„akzeptiert“ und „nicht akzeptiert“ in der einen oder anderen Richtung überschritten wird. Kommt es zwischen den beiden Verkostungen zusätzlich zu einer Veränderung des Weins, weil zum Beispiel ein Bockser stärker oder schwächer wird, erhalten wir einen dritten Faktor, den man mit „Wein x Zeit“ umschreibt.

Die Anwendung der Varianzanalyse auf sensorische Daten lässt erkennen, inwiefern das Prüfungsergebnis abgesichert ist. So wird vermieden, dass scheinbare, nur auf die Unzuverlässigkeit der Prüfer zurückzuführende Unterschiede zwischen Weinen vor schnell zu Gesetzmäßigkeiten hochgerechnet werden. Man kann ohne Übertreibung sagen, dass die **Sensorik mit der Statistik lebt oder stirbt**.

### Schwierige Frage nach Qualität

Grundsätzlich anderen Gesetzen als denen der rationalen Sensorik gehorcht die hedonische Antwort von Kostern auf einen gegebenen Wein. Unter Hedonismus fasst man in der Sensorik die Aspekte der subjektiven Vorlieben und persönlichen Qualitätsvorstellungen zusammen. Kurz, was empfindet man als gut oder schlecht?

Kulturelles Umfeld und persönliche Erfahrungen eines Menschen bestimmen weitgehend seine hedonischen Reaktionen wie Wohlempfinden, Akzeptanz oder Ablehnung. Ob-

wohl gewisse Präferenzmuster beim Menschen angeboren sein mögen, unterliegen sie ständigen Veränderungen durch Ernährungs- und Stoffwechselprozesse, Alter, Gewöhnung und Erfahrung. Die Vertrautheit mit dem Produkt, Intensität und Komplexität der von ihm vermittelten Stimuli, aber auch Ermüdung nach längerem Konsum haben Einfluss auf das, was man als gut oder schlecht empfindet. Aus diesem Grund kann die hedonische Antwort eines Prüfers kein Maßstab für seine fachliche Qualifikation sein.

*Wer Spitzenwinzer werden will, muss Weinpäpste für sich und seine Weine emotional als auch finanziell vereinnahmen.*

Trotzdem müssen Prüfer ein Urteil über Weinqualität abgeben. Die Aussagekraft eines solchen Urteils ist nicht durch objektive Tests zu ermitteln, **denn Qualität ist und bleibt eine subjektive Auffassung**. Man kann von einem guten Prüfer nur verlangen, dass er in der Lage ist, sein Qualitätsurteil zuverlässig zu reproduzieren. Es zählt nicht zu den Anforderungen an ihn, Weine bestimmter Stilrichtungen zu mögen, abzulehnen oder vorzuziehen. Die Frage der Prüferqualifikation bleibt von den persönlichen Präferenzen unberührt.

Fachlich geschulte Prüfer sollen auch in der Lage sein, objektive Weinfehler eindeutig zu identifizieren. Ob sie fehlerbehaftete Weine unbedingt ablehnen, ist eine andere Frage. Die Praxis gibt darauf eine eindeutige Antwort. In der Tat vermag der richtige Meinungsbildner am richtigen Ort und zum rechten Moment wahre Wunder zu vollbringen. Etwa wenn es darum geht,

einen in jeder Qualitätsweinprüfung abgelehnten Bockser mit den rhetorischen Erfolgsformeln von Terroir, Komplexität oder Mineralik vor einem völlig verunsicherten Publikum zu einem individuellen Weinstil zu mystifizieren. Wer mag es da dem einen oder anderen Winzer noch verübeln, wenn er seinen UTA mag?

### Wettbewerb und Prämierungen

Der Qualitätsbegriff ist von einer Vielzahl persönlicher Faktoren wie Erziehung, Gewohnheit und kulturellem Umfeld abhängig, die nicht konsensfähig sind. Sie führen bei Fachleuten zu Bewertungs-

mustern, die mit den hedonischen Maßstäben ungeschulter Verbraucher kaum vergleichbar sind. Deshalb setzt die professionelle Marktforschung in der Lebensmittelindustrie nie Experten zur Untersuchung hedonischer Aspekte ein. Es ist eines der **Grundübel der Weinbranche**, dass zu oft Erzeuger unter sich ihre Produkte bewerten und sich gegenseitig beibräuchern.

Grundsätzlich kann die Qualitätsbewertung eines Weins nur die tendenzielle Meinung einer Prüfergruppe wiedergeben. Sie erlaubt nie, die Qualitätseinstufung des Weins durch einen einzelnen Koster vorherzusagen. Damit das System überhaupt funktioniert, ist zwischen den Prüfern eine Art von Mindestkonsens über Qualität notwendig. Sie müssen eine annähernd gemeinsame Erwartungshaltung oder Idealvorstellung von dem zu prüfenden Wein haben. Dies ist im engeren geographischen Umfeld oder Anbaugebiet meist gegeben. Kritiker werfen den auf regionaler Ebene durchgeführten Qualitätsbewertungen vor, dass sie **einen gebietstypischen Mainstream vorgeben**, innerhalb dessen individuelle Variationen keinen Platz finden und mit Nichtbeachtung abgestraft werden. Sie haben indessen den Vorteil, dass sie hervorragend zur Beurteilung und Festschreibung eines lokalen Weintyps geeignet sind. Selbst zuverlässige Prüfer versagen oft, wenn Weine zu prüfen sind, deren Typ ihnen wenig geläufig ist.

Vor- und Nachteile kehren sich um in dem Maß, wie sich der Bewertungskontext geographisch öffnet bis

**Die Aussagekraft sensorischer Ergebnisse hängt von der Qualifikation der Prüfer ab – jeder Prüfer hat seine Stärken und Schwächen, etwa beim Erkennen von einzelnen Aromen**





hin zu den internationalen Wettbewerben. Je breiter und internationaler das Spektrum der Weine und ihrer Prüfer, desto weniger Berücksichtigung finden regionale Weinstile oder Sorten zu Gunsten eines standardisierten globalen Geschmacks. Erst in jüngerer Zeit wurden internationale, rebsortenspezifische Wettbewerbe ins Leben gerufen, mit dem Vorteil, dass Weine mit ähnlichen Charakteristika eine wenigstens halbwegs vergleichbare Beurteilungsbasis schaffen. Eine solche ist zwischen einem kalifornischen Chardonnay und einem Grünen Veltliner kaum gegeben.

Da mit **Weinwettbewerben sehr viel Geld umgesetzt** wird, schießen sie als eigenständige Industrie überall wie Pilze aus dem Boden. Die Auswahl der Prüfer erfolgt fast immer nach rein subjektiven Kriterien, wobei persönliche Beziehungen, die Zugehörigkeit zu einem Verband oder Institut als auch die Position in der beruflichen Hierarchie die entscheidende Rolle spielen. Es ist nicht üblich, von den Prüfern einen Qualifikationsnachweis zu fordern oder ihr persönliches Bewertungsverhalten zu berücksichtigen, wie es selbst bei der viel verschmähten Qualitätsweinprüfung praktiziert wird. Verständlich erscheint daher die immer wiederkehrende Kritik, wenn grob fehlerhafte Weine oder solche aus unreifen Trauben eine Goldmedaille erhalten. Auf jeden Fall sollte man als Erzeuger vor einer Teilnahme entscheiden, in welchem Kontext, nach welchen Kriterien und von welchen Prüfern man seinen Wein bewertet haben will. Entsprechende Informationen sind gezielt zu erfragen.

### Vorsicht mit Punktzahlen

In allen Wettbewerben wird die **Qualität in Form einer Punktzahl ausgedrückt**. Die Verwendung von Punkten kommt dem natürlichen Bedürf-

nis des Menschen entgegen, komplexe Zusammenhänge auf einfache Zahlenwerte zu reduzieren. Skalen

*Da mit Weinwettbewerben sehr viel Geld umgesetzt wird, schießen sie als eigenständige Industrie überall wie Pilze aus dem Boden.*

von 5, 20 und 100 Punkten haben alle ihre Vor- und Nachteile. Entscheidend ist, wie der Koster mit dem Gebrauch der Skala vertraut ist.

Innerhalb dieser Skalen sind die Intervalle von Punkt zu Punkt theoretisch gleich groß, sie werden jedoch aus psychologischen Gründen unterschiedlich gehandhabt. Die mittleren Zahlen der Skala werden bevorzugt und die Skalenden kaum benutzt. Oft existiert eine ungeschriebene psychologische Grenze, die im allgemeinen Sprachgebrauch mit „über oder unter dem Strich“ bezeichnet wird. Da sich die Unterschiedsgrößen innerhalb der Skala rechnerisch kaum fassen lassen, ist das arithmetische Mittel mit Vorsicht zu interpretieren. Trotzdem werden auf Grund derartiger erhaltener Gesamtpunktzahlen Prämierungsklassen aufgestellt und Handelspreise festgelegt.

Die unterschiedliche Verwendung von Punktskalen ist nur eine Frage, welche über ein vorgelagertes und viel größeres Problem hinwegtäuschen kann: das der Definition von Qualität. Da Qualität nicht objektiv messbar ist, kann sie kein Gegenstand von Abhandlungen zur Sensorik werden, sondern bestenfalls zum Thema nie enden wollender Streitfragen.

### Qualitätsgarantie durch Weinpöpste?

An genau diesen Streitfragen versuchen sich einzelne Weinkritiker zu profilieren, auch Weinpöpste genannt. Sie ernennen sich selbst zu dem, was sie sind. Niemand verlangt eine Überprüfung ihrer sensorischen Kompetenz. Sie legen die Hackordnung in der Winzerzunft fest und klassifizieren die Wertigkeit von Weinen nach ihren eigenen unantastbaren Kriterien. Weine werden nicht zu internationalen Spitzenweinen, weil die Qualität plötzlich objektivierbar geworden ist, sondern weil sie von einschlägigen Meinungsbildnern

dazu gemacht werden. Obwohl Einzelbewertungen von zweifelhaftem Wert sind, wird Qualität durch die Autorität einer einzelnen Person legitimiert. Der Punktefetischismus dient als Instrument. Wer Spitzenwinzer werden will, muss Weinpöpste für sich

und seine Weine emotional als auch finanziell vereinnahmen.

Qualität kann weder gepachtet noch genormt werden. Sie ist ein Produkt von Zeitgeist, kulturellem Umfeld, persönlicher Gewohnheit – und damit letztendlich auch ein kulturelles Modell. Die Anmaßung, aus persönlichen Präferenzen absolute Qualitätsnormen ableiten zu wollen, entspricht einer Machtergreifung der besonderen Art, die sich von Selbstherrlichkeit nährt. Die nationalen Önologenverbände haben bis jetzt bemerkenswert wenig unternommen, diesem Treiben entgegenzuwirken. Sie wären als Erste berufen, die Unterschiede zwischen sensorisch objektiven Kriterien und persönlichen Vorlieben aufzuzeigen. Doch im Gegenteil, die **Globalisierung macht nirgends größere Fortschritte als in der Normierung des Weingeschmacks**. Heerscharen fliegender „winemaker“ als vermeintliche Garanten hoher Parker-Punkte sind bestrebt, dem universalen Geschmack zu weltweiter Geltung zu verhelfen. Mit beispielloser Unterwürfigkeit akzeptiert die halbe Weinwelt willens- und kritiklos, dass der Qualitätsbegriff durch selbst ernannte Meinungsbildner zu Lasten eines demokratisch legitimierten Qualitätsverständnisses monopolisiert wird – wobei Individualität auf der Strecke bleibt und gewachsene Weinkulturen diffamiert werden.

### Zusammenfassung

Die Qualifikation von Kestern kann durch die Bestimmung ihres Diskriminiervermögens, ihrer Sensibilität und ihrer Reproduzierbarkeit objektiv ermittelt werden. Mittels mathematisch-statistischer Verfahren wird die Aussagekraft sensorischer Prüfergebnisse beurteilt. Qualität als kulturelles Modell ist nicht zu standardisieren. Entsprechend kritisch sind Punktzahlen von Weinwettbewerben und Weinpöpsten zu interpretieren.

#### Der Autor

Volker Schneider,  
Schneider-Oenologie,  
Am Entenbach 5, 55411 Bingen/Deutschland,  
Tel.: +49(0)6721/18 27 64, Fax: 18 27 65,  
www.Schneider-Oenologie.com

